UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Prostate Lesion Segmentation with Convolutional Neural Networks

Luís Miguel Teixeira Faria Venâncio

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em Engenharia Clínica e Instrumentação Médica

Dissertação orientada por:
Dr. Nuno André da Silva
Dra. Raquel Conceição

2020

Acknowledgements

First of all, I would like to thank Doctor Nuno André da Silva, for accepting me in the Learning Health group at Hospital da Luz Lisboa and for all the support and guidance during this academic year. It was a great pleasure being part of such an important medical institution in Portugal that is always at the forefront of clinical research.

I would also like to express my sincere appreciation to Professor Raquel Cruz Conceição, from Faculty of Sciences of the University of Lisbon, for all the support, regular and constructive feedback and motivation throughout this project.

Special mention goes to Dr Adalgisa Guerra, from the medical imaging department of Hospital da Luz Lisboa. I am grateful for all the hours that she spent with me analysing magnetic resonance images and all the clinical insights that were truly relevant for this project development. Without her contribution, this dissertation would not be possible. Another special mention goes to André Barbosa, from Siemens, who was always available to help and teach how to manage clinical software necessary for this project.

To all my close friends, I am truly grateful for always being available along the way, either for friendship or academic purposes. This five-year journey would not have been the same without them.

Last, but not least, I owe more than thanks to all my family, but specially to my grandparents and my mother. I am sincerely grateful for the unconditionally support, guidance, motivation and encouragement throughout my life. I would not have reached far without them.

Resumo

O cancro da próstata é o segundo tipo de cancro não cutâneo com maior incidência nos homens em todo o mundo, a seguir ao cancro do pulmão. Em Portugal, de acordo com a *Associação Portuguesa de Urologia*, esta doença representa, aproximadamente, 3,5% de todas as mortes nacionais, assim como 10% das mortes relacionadas com cancro. Para além destes dados, o *Global Cancer Observatory*, estima que a probabilidade de um homem ocidental ser diagnosticado ao longo da sua vida com cancro da próstata é de 8,1%. As causas diretas que levam ao aparecimento deste tipo de cancro ainda não estão totalmente clarificadas, no entanto, os hábitos alimentares, o estilo de vida e o ambiente em redor desempenham um fator preponderante no desencadeamento desta patologia. A deteção inicial deste cancro ocorre, normalmente, através de exames retais de rotina, ou através de alterações significativas do antigénio prostático específico detetáveis em análises ao sangue. De seguida, para confirmação e localização do possível tumor, podem ser adotados três procedimentos: ecografia transrectal, colheita de uma biópsia local ou análise de imagem prostática através de ressonância magnética. Por ser o procedimento menos invasivo, a ressonância magnética é a ferramenta mais utilizada para deteção e localização de lesões na próstata.

No Hospital da Luz de Lisboa, a análise de imagens provenientes de ressonância magnética multi-paramétrica é o procedimento padrão para a localização de lesões prostáticas. Neste exame, geralmente, são adquiridas três sequências em T2, uma em cada um dos planos axial, coronal e sagital, duas sequências com difusão e uma sequência em T1. Cada exame demora, aproximadamente, 45 minutos a ser analisado corretamente pelo radiologista. Após a análise, é atribuída uma classificação ao estado do paciente, de T1 a T4, sendo que até T2 o tumor ainda se encontra exclusivamente no interior da próstata e em T4 apresenta os maiores índices de disseminação em redor da próstata. Esta classificação é preponderante para o planeamento da cirurgia de remoção do tumor. Nesta avaliação, é normalmente identificada a lesão ”índex” da próstata, que corresponde à lesão com maior índice cancerígena e, por isso, a mais visível. No entanto, podem em certos casos existir lesões de menor dimensão ou de menor relevância, lesões ”não-índex”, que em determinadas circunstâncias levam à alteração da classificação do estado do paciente. Este tipo de lesões, por vezes, não é facilmente localizado e o procedimento cirúrgico resultante acaba por não ser o mais indicado e gerar, no futuro, reincidências. Até T2, a prostatectomia deve ser realizada com o intuito de remover apenas a lesão ou a próstata por completo, no entanto, em T3 e em T4, a abordagem deve ser um pouco mais severa, sendo necessário também remover camadas celulares fora da próstata como margem de segurança para evitar uma reincidência.

A introdução de algoritmos de inteligência artifical no ramo da medicina, com o propósito de realizar tarefas como segmentação, classificação e deteção de artefactos em imagens digitais, tem sido cada vez mais preponderante na evolução tecnológica da saúde. No panorama geral da medicina, os métodos de avaliação automatizada permitem executar tarefas com maior rapidez, precisão e assertividade face à capacidade humana, sendo possível explorar numa imagem, por exemplo, texturas, formas, estruturas e até mesmo orientações nucleares de certos artefactos. Relativamente ao cancro da próstata, para além de algoritmos que visam auxiliar as avaliações promovidas pela anatomia patológica, o grande foco centra-se em melhorar os métodos de análise de imagem de ressonância, por forma a tornar os diagnósticos mais precisos. Assim sendo, a criação de algoritmos que permitam a segmentação das lesões prostáticas, assim

como respetiva ponderação da classificação do estado do paciente, revela-se como a tarefa principal na evolução do diagnóstico do cancro da próstata.

Desta forma, com o objetivo de otimizar a deteção e localização das lesões prostáticas, esta dissertação apresenta um conjunto de algoritmos que visam a segmentação de lesões da próstata em imagens de ressonância magnética. O projeto foi desenvolvido no centro de formação e investigação Learning Health, no Hospital da Luz de Lisboa, e apresenta duas etapas principais: a criação do modelo de segmentação da próstata e a elaboração do modelo de segmentação das lesões prostáticas.

Na fase inicial desta dissertação, a criação de um modelo que segmentasse a zona da próstata, por forma a aumentar, posteriormente, a área de deteção das lesões, foi identificado como o primeiro passo. Com base em modelos de *deep learning*, mais especificamente através de *convolutional neural networks*, foi desenvolvida uma arquitetura para o propósito anteriormente descrito. Esta arquitetura, baseada numa rede já previamente construída, a *U-Net*, apresenta características específicas que permitem a entrada de imagens de ressonância magnética da próstata, *slice* a *slice*, a gestão da informação que essas imagens apresentam e, por fim, a criação de máscaras binárias da zona da próstata consoante a *slice* de entrada. Com as máscaras da zona prostática, foi possível delinear um contorno e promover uma sub-seleção dessa zona na imagem original, criando volumes onde a área de deteção das lesões da próstata é isolada.

Na segunda fase deste projeto, foi criado um modelo para segmentar diretamente as lesões da próstata. Para tal, foram utilizadas as imagens adquiridas após a primeira parte do projeto, assim como a rede identificada para localizar a próstata. Contudo, esta arquitetura sofreu alterações estruturais, por forma a otimizar o rendimento do modelo. Ao contrário da rede anterior, esta arquitetura permite a entrada de duas imagens na mesma instância, a original T2 e a respetiva original ADC. No final, o *output* é, igualmente, uma máscara binária, desta vez localizando as lesões da próstata em imagens de ressonância.

Em ambos os modelos, foram utilizadas como imagens de *input*, casos de ressoância magnética adquiridos no Hospital da Luz de Lisboa. Para este processo final, foi necessário segmentar manualmente tanto a próstata, como as respetivas lesões, nas imagens do hospital. Para tal, utilizou-se um software hospitalar, o *Multi-Parametric Analysis*, que permite o registo das imagens originais e a elaboração das máscaras manualmente. Este processo de identificação e elaboração manual das máscaras da próstata e das lesões foi realizado por uma radiologista do Hospital da Luz de Lisboa, a Dra. Adalgisa Guerra.

O modelo desenvolvido na primeira etapa, para a segmentação da próstata, apresentou um valor de *Dice Similarity Coefficient*, a principal métrica de avaliação em projetos de segmentação, de 0,88. Este valor é semelhante aos valores de referência destacados no *state of the art*. Após a conclusão desta etapa, criaram-se cinco modelos para segmentar as lesões da próstata, sendo que o modelo que apresentou melhores resultados foi o que tinha como *input* as imagens ampliadas da próstata em T2 e ADC e as respetivas máscaras das lesões criadas em imagens T2. O resultado final deste modelo em termos de *Dice Similarity Coefficient* foi de 0,76, *Hausdorff Distance* de 20,2 mm e *Mean Square Distance* de 2,1 mm. Este resultado realça o impacto que a informação combinada de duas sequências consegue ter no processo de segmentação de lesões da próstata.

Concluindo, a medicina, em consonância com as restantes áreas da sociedade, está a evoluir e a inteligência artificial terá um papel preponderante nessa transição. Neste caso, esta dissertação pretende otimizar a metodologia atual utilizada num hospital local, conferindo aos profissionais de saúde cada vez mais e melhores condições para realizarem as suas tarefas.

Palavras-Chave: Cancro; *Deep Learning*; Ressonância Magnética; Segmentação, *U-Net*.

Abstract

Prostate cancer is the second most commonly diagnosed non-cutaneous cancer in men, in many parts of the western world, and is a major cause of cancer-related death. In Europe, according to *Global Cancer Observatory*, the risk of a Western man being diagnosed during his life with prostate cancer is 8.1%. In Portugal, according to *Associação Portuguesa de Urologia*, it is estimated that prostate cancer represents 3.5% of all deaths and 10% of all the cancer deaths in the country. The causes are not totally clear, but it is thought that food habits, lifestyle or environment have an important role in cancer prevalence. As can be seen, prostate cancer unfortunately has a significant impact in our society, and so it is necessary to improve our methods of diagnosis.

Currently, the most commonly used techniques to detect whether there is an abnormality in the prostate are the Digital Rectal Exam (DRE) and the analysis of Prostate-Specific Antigen (PSA). Both tests are important to identify prostate cancer at an early stage. After these procedures, it is important to determine if the detected abnormality in the prostate is a cancer and identify its localisation. To perform these, there are three main procedures: transrectal ultrasound, collecting a sample of prostate tissue and Magnetic Resonance Imaging (MRI). The first two procedures are invasive, and so, MRI, especially multi-parametric MRI (mp-MRI), where multiple images are acquired, is the most useful tool for detecting and localising prostate lesions.

At Hospital da Luz de Lisboa, each exam takes approximately 45 minutes to be analysed in order to detect and localise the most relevant prostate lesions. Index lesions are defined as the lesions with the highest cancer suspicion score based on initial mp-MRI of a patient, irrespective of size. These lesions have an impact on patients outcomes by dictating the stage of the cancer. The stage can be defined as T1 (tumour too small to be seen on a scan), a T2 (tumour found only in the prostate), T3 (tumour has spread out of the prostate on one side and into the tissue just outside the prostate) or T4 (tumour has disseminated into nearby structures other than the seminal vesicles). The stage classification is relevant because it is through this evaluation that the surgeon knows whether to only remove the prostate (prostatectomy) or the prostate plus a safety margin, considering the intensity and the localisation of the lesion.

Conversely, identifying non-index lesions is complicated and a significantly time consuming task for radiologists. However, if a non-index lesion is near the border of the prostate, the stage of the patient can be reconsidered and sometimes changed from a T2 to a T3 or T4. This is relevant because, at a T3 or T4 stage, both prostate and margin should be removed in order to avoid a relapse of the disease.

This dissertation is then focused on the creation of an automatic model that is able to segment prostate lesions, by exploiting deep learning models. The project was developed at the research centre *Learning Health*, from Hospital da Luz de Lisboa, and is divided in two main sections: the creation of an algorithm to segment the prostate zone and the building of a model capable of segmenting the prostatic lesions.

At first, cropping the region around the prostate was performed, in order to restrict the lesion detection area to that of the gland. After that, a model based on U-Net architecture was developed, receiving as input each original slice of the MRI and getting as output a binary mask of the prostate. Thus, outlining and cropping the images were possible, creating a second dataset with the region of interest.

The second part of the project consisted of the creation of a different model, this time to segment

the prostate lesions. It was also based on a U-Net architecture, but it has some structural changes in comparison with the previous model. This model is capable of receiving an ADC image, as well as a T2 image. The output results are equally binary masks with the proper segmentation of the prostatic lesions.

The developed model for prostate region segmentation demonstrated a performance of 0.88, as measured by Dice Similarity Coefficient, the principal metric used in image segmentation projects. This performance is similar to that reported in the state-of-the-art. After this task, five models were trained to segment prostate lesions and the model that proved better results presented a mean Dice Similarity Coefficient of 0.76, an Hausdorff Distance of 20.2 mm and a Mean Square Distance of 2.1 mm. This model received as input T2 and ADC cropped images with the respective masks, annotated in T2 images, emphasising the impact that both sequences can have in prostate lesion segmentation.

This document describes in detail all the steps completed, as well as the final results. In conclusion, medicine, as well as other sectors of our society, is evolving, and artificial intelligence has a relevant role in this transition. The project presented in this dissertation helped to further optimise the current methodology at a local hospital, an important step in improving the tools with which healthcare professionals perform their duties.

Keywords: Prostate Cancer; Deep Learning; Magnetic Resonance Imaging; Segmentation, U-Net.

List of Figures

List of Tables

Acronyms

AdaM	Adaptive Moment
ADC	Apparent Diffusion Coefficient
AI	Artificial Intelligence
AUC	Area Under the Curve
BCR	BioChemical Recurrence
CNN	Convolutional Neural Network
DRE	Digital Rectal Exam
DSC	Dice Similarity Coefficient
DWI	Diffusion Weighted-Image
HD	Hausdorff Distance
IoU	Intersection over Union
LR	Learning Rate
MAE	Mean Absolute Error
MRI	Magnetic Resonance Imaging
MSD	Mean Square Distance
MSE	Mean Square Error
PI-RADS	Prostate Imaging Reporting and Data System
PSA	Prostate-Specific Antigen
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
SVI	Seminal Vesicle Involvement
TRUS	Transrectal Ultrasound
T2WI	T2 Weighted-Image

Table of Contents

1. Introduction

In this dissertation, an approach to segmentation of prostate lesions, in magnetic resonance images, using Convolutional Neural Network (CNN)s is presented. This first chapter contextualises the health issue under study, as well as describes the motivation for the project, major contributions and, finally, the dissertation outline.

1.1 Context and Motivation

Prostate cancer is a pathology with significant incidence in our society. Currently, it is considered the second most commonly diagnosed non-cutaneous cancer in men, in many parts of the Western world, and is a major cause of cancer-related death. There are several approaches to detect whether there is an abnormality in the prostate. Magnetic Resonance Imaging (MRI) is the most applied imaging modality since it is minimally invasive and provides a good soft tissue contrast, crucial to localise prostate lesions.

At Hospital da Luz Lisboa, multi-parametric MRI - an approach where multiple images are acquired - is used to detect prostate lesions. Due to the large number of images, analysing these exams is time consuming. For instance, each exam takes approximately 45 minutes to be analysed, including detecting and staging the lesions. The stage can be defined as T1 (tumour too small to be seen on a scan), T2 (tumour found only in the prostate), T3 (tumour has spread out of the prostate on one side and into the tissue just outside the prostate) or T4 (tumour has disseminated into nearby structures other than the seminal vesicles). This classification is relevant because it allows the surgeon to know whether to only remove the prostate (prostatectomy) or the prostate plus a safety margin, considering the intensity and the localisation of the lesion. Taking into account the repetitive nature of the task, manually detecting all lesions might not be a reasonable approach. In addition, missing the detection of a lesion can also compromise the stage classification and, consequently, the surgery procedure.

Since the rise of Artificial Intelligence (AI) within the medical imaging field, a revolutionary stage in segmentation, classification or even detection of artefacts in images is undergoing. These methods allow quicker execution of certain repetitive tasks, with more precision and assertiveness, optimising the human work.

Therefore, in order to optimise the current methodology at Hospital da Luz Lisboa and to provide healthcare professionals with more and better resources to perform their work, this dissertation was idealised, developed and it is now presented in this manuscript which describes the creation of automatic algorithms with CNN, capable of segmenting prostate lesions.

1.2 Objectives

The overall objective of this dissertation is to develop an AI based algorithm, using CNNs, to identify and segment prostate lesions in magnetic resonance images. Furthermore, there are other secondary objectives, including:

1. Comprehension in greater detail of prostate cancer and its national and global incidence.
2. Analysis of the current literature in order to identify the state of the art of automatic algorithms developed to optimise prostate cancer segmentation.
3. Understanding the main issues related to the creation of CNNs applied to medical imaging.
4. Set up a database of prostate magnetic resonance images, as well as the respective manual segmentation of the lesions.
5. Design of neural networks to be applied to the above-mentioned database.
6. Comparison of the impact that image contrast has.
7. Evaluation and discussion of the respective results.

1.3 Major Contributions

The major contribution of this dissertation is to explore potential optimisations to the process used by healthcare professionals to segment prostate lesions, especially those of Hospital da Luz Lisboa. So far, the delineation and identification of the lesions are manually completed, leading to excessive time commitment by the radiologists. The proposed model is then important to reduce the time each exam takes to be evaluated, allowing the optimisation of clinician's time. Furthermore, different models, with specific architectures, were also applied and tested in magnetic resonance images in order to segment prostate lesions.

1.4 Outline

In order to achieve the main goals of this dissertation, this project is divided into seven Chapters. The current chapter provides the introduction to the dissertation theme by describing the context and motivation that lead to its development, as well as, principal objectives, main contributions and general organisation of this document.

Chapter 2 presents the description of the basic concepts of this work. It includes prostate cancer background and common procedures for treatment and diagnosis. Likewise, a comprehensive explanation of deep learning fundamental concepts regarding CNNs and their development are also provided. Chapter 3 reports the state of the art of deep learning for image segmentation by describing the most relevant network architectures and its direct influence in medical imaging. The principal methodology applied to

this project, as well as the material used for its development are detailed in Chapter 4. In Chapter 5, the results of the work are presented, while in Chapter 6 those results are discussed and validated. Finally, Chapter 7 describes a brief overview of the main findings and its conclusions, as well as future work perspectives.

2. Basic Concepts

This chapter presents the basic concepts and resources that support the conceptualisation of a neural network to segment prostate lesions by describing in greater detail prostate cancer (2.1) and also deep learning models (2.2).

2.1 Prostate Cancer

This section describes prostate cancer. An introduction of prostate anatomy is presented, followed by prostate cancer incidence, diagnosis and treatment. The remaining subsection describes the main procedures to detect prostate lesions adopted at Hospital da Luz, where this dissertation was developed.

2.1.1 Prostate Anatomy

The adult prostate is defined as an elastic gland, found only in men, which has the shape of an inverted pyramid, located at the base of the bladder (Figure 2.1). The main function of this organ is to produce a white fluid, which is one of the components of semen. Anatomically, the prostate is traversed by the prostatic urethra and the paired ejaculatory duct, the latter emptying distally in the verumontanum, which is a small mound of tissue situated in the distal prostatic urethra. The verumontanum contains a small central cavity that communicates with the urethra, called uritcle [1]. The Denonvilliers' fascia - a thin, filmy layer of connective tissue - is responsible for separating the seminal vesicles and the prostate from the rectum posteriorly [2]. In terms of dimension, the average prostate has 3.4cm in length, 4.4cm in width and 2.6cm in thickness, while the weight is around 20g [1].

The prostate has two main lobes, left and right, and can be divided into five zones: peripheral zone, transitional zone, central zone, verumontanum and urethra (Figure 2.2). The transitional zone can be identified in the central part of the prostate and this is where the benign growth usually occurs in most men above the age of 50. Conversely, the peripheral zone is found next to the rectum, on both sides of the urethra, and this is normally where the relevant prostate cancers develop [2]. These two regions are extremely important to evaluate prostate cancer incidence.

The seminal vesicles are posteriorly and superiorly located at the base of the prostate. They are also removed during prostatectomy because they are a likely location for cancer cells to spread. Seminal vesicles undergo confluence with vas deferens on each side to form the ejaculatory duct complex, which is the set of two ejaculatory ducts, along with a second loose stroma. The Seminal Vesicle Involvement (SVI), which consists of the presence of prostate cancer in the areolar connective tissue around the seminal vesicles and outside the prostate, is one of the most relevant predictors of cancer progression [4; 5].

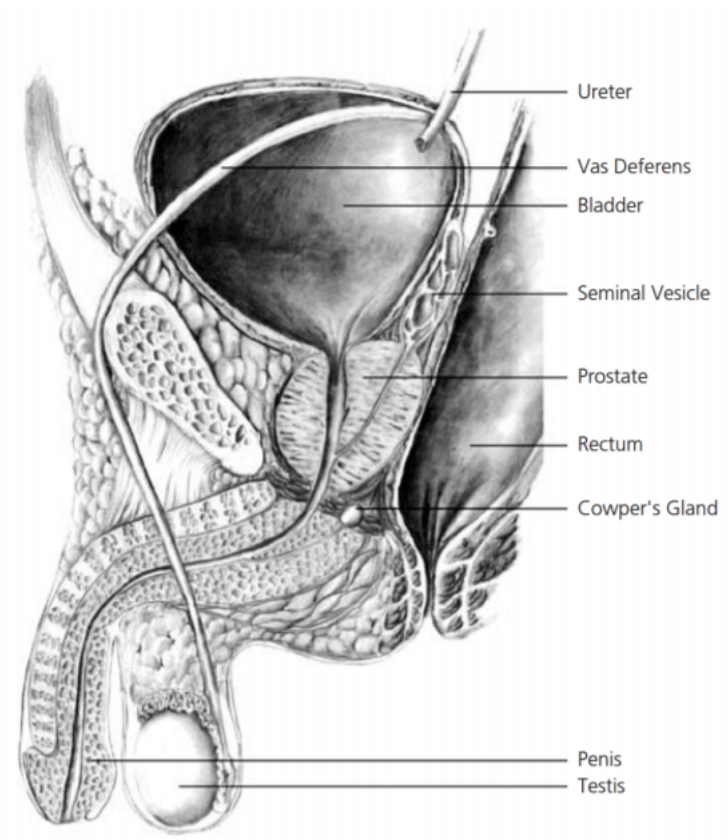

Figure 2.1: Sagittal section of the male pelvis. Adapted from [3].

The prostatic capsule is composed of fibromuscular stroma which disappears towards the apex of the gland. Although the term "capsule" is regularly used and found in the common medical literature, there are no certainties about the real presence of an actual capsule [6; 7].

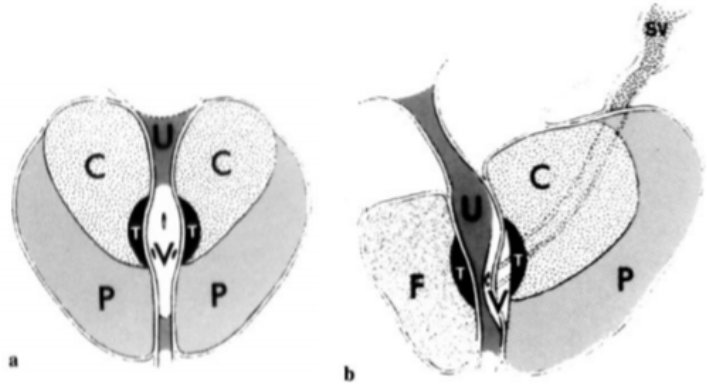

Figure 2.2: **a** Coronal and **b** sagittal diagrams of the prostate showing zonal anatomy. The transition zone comprises only 5%–10% of the glandular tissue in the young male. The central zone forms part of the base of the prostate and it is traversed by the ejaculatory ducts. T, transitional zone; P, peripheral zone; C, central zone; V, verumontanum; U, urethra; F, anterior fibrous zone. Adapted from [1].

2.1.2 Prostate Cancer Incidence

Prostate cancer is the most commonly diagnosed non-cutaneous cancer in men, in many parts of the Western world, and is a major cause of cancer-related death. In Europe, according to *Global Cancer Observatory*, 449,761 men were diagnosed with prostate cancer and 107,315 died due to this disease in 2018 [8]. Furthermore, the risk of a Western man being diagnosed during his life with prostate cancer is 8.1%. In Portugal, according to *Associação Portuguesa de Urologia* [9], it is estimated that prostate cancer has an incidence of 82 cases per 100,000 inhabitants and a mortality rate of 40% in diagnosed cases. It also represents 3.5% of all deaths in Portugal and 10% of all the cancer deaths in the country. The causes are not completely clear, but it is thought that food habits, lifestyle or environment play an important role in this situation [10]. Unfortunately, this type of cancer has a negative impact in our society and it is necessary to improve methods to provide better diagnosis.

Prostate cancer growth can be distinguished by two main forms of evolution: non-clinically-significant tumours and clinically-significant tumours. The first type comprises around 85% of all prostate cancers and is generally focused on the prostate gland. In these cases, frequent monitoring is enough and there is no need for specific treatment. However, clinically-significant tumours progress rapidly, metastasising from the prostate gland to other organs and bones easily. To tackle this problem, certain procedures are adopted, such as radical surgery, removing the cancerous region, or radiation therapy, to eradicate cancer cells [11; 12].

2.1.3 Prostate Cancer Diagnosis

Prostate cancer can be considered a clinically silent pathology [13; 14]. In the last few decades, there has been a significant difficulty to detect this type of cancer during its early stages, and as a result, improving the methods to identify prostate lesion became the main focus for researchers in this area. The introduction and widespread of screening with serum Prostate-Specific Antigen (PSA) blood test has considerably increased the number of men diagnosed with prostate cancer. This growth is important because it allows clinicians to identify more assertive strategies to fight cancer while it is not spread throughout the body.

Currently, the most commonly used techniques to detect whether there is an abnormality in the prostate are the (Digital Rectal Exam (DRE)) and the analysis of PSA. The DRE allows the physician to detect palpable abnormalities, such as modules, hardness and asymmetry, mainly in the lateral and posterior regions of the prostate gland, where most of the relevant cancer lesions develop [15]. Conversely, PSA is simply a test blood with the purpose of measuring the PSA protein level, and a concentration of more than 4ng/mL (nanograms per milliliter) is potentially suggestive of prostate cancer presence. However, this PSA test is not the most reliable option, since prostatic infections or even prostatic hyperplasia can also increase the level of this protein [16]. Therefore, both tests are important to detect whether the prostate is completely healthy and if not, more advanced strategies should be considered to identify the problem.

After the previous procedures, if a problem is detected, it is important to determine if the respective abnormality in the prostate is cancer and its location. This way, MRI, especially multi-parametric

MRI (mp-MRI), where multiple images are acquired, is the most useful tool for detecting and localising prostate lesions [17]. Mp-MRI consists of combining the morphological and standard assessment of T2 Weighted-Image (T2WI), or T1, with different functional MRI techniques, such as Dynamic Contrast-Enhanced (DCE) perfusion imaging, Diffusion Weighted-Image (DWI) and MR Spectroscopic Imaging (MRSI). In particular, T2WI and DWI have shown promising results in detection, localisation and staging prostate cancer [18]. One of the most important advantages of the mp-MRI technique is the excellent sensitivity for prostate lesions with Gleason score higher than 7, more specifically lesions with a volume of at least 0.5 mL [19]. The Gleason score is the sum of the two most prevalent pathological tumour patterns, and ranges from 2 to 10. Gleason scores of seven or greater are generally considered indicative of clinically-significant prostate cancer. In order to standardise the acquisition, interpretation and reporting of magnetic resonance prostate images for prostate cancer diagnosis, the first version of the Prostate Imaging Reporting and Data System (PI-RADS) was introduced in 2012.

To achieve a definitive diagnosis, a biopsy is performed, in which a sample is collected in a small intervention and examined under a microscope by a pathologist, who is responsible for the evaluation of the histological samples and final decision if prostate cancer is present.

2.1.4 Prostate Cancer Treatment

After combining PSA, DRE and MRI results, along with pathological biopsy analysis, the stage and nature of the cancer are determined. As previously mentioned, there are two main approaches to treat this disease: radiation therapy and prostatectomy. The impact of prostate cancer treatments is generally present for many years - mostly by urinary incontinence - and it is directly related to the proper treatment selection, enhancing how precise and justified the choice should be [3].

When the disease is already in advanced stages, with extended involvement of prostate surrounding areas, a directly invasive approach may be inappropriate, since it would unnecessarily endanger the patient while burdening healthcare resources. Therefore, removing all the cancerous regions when there are significant metastases, is not the best option. In these cases, radiation therapy is the main treatment, because it can target all the considered areas, eradicating cancer cells [3].

Conversely, surgical prostatectomy is generally applied when there is a localised prostate cancer, where the lesion is directly confined inside the prostate and does not affect the surrounding areas, such as seminal vesicles, lymph nodes or distant site. This is the standard first-line treatment with curative intent [20; 21]. The main advantage of this approach is that when the prostate is surgically removed, there is a significant reduction of metastasis risk [22]. Since this is the most applied method, over the past few years there was a considerable improvement in this type of treatment, especially with the growth of surgical robotics in the health sector. One good example is the Robotic-Assisted Laparoscopic Prostatectomy (RALP), Figure 2.3, a procedure which allows the clinician to perform complex surgeries with both skill and minimal fatigue, and also reduces the operative time, blood loss and hospitalisation time for being less invasive than the previous prostatectomy procedures [23; 24].

The improvements in this field are notable, however, there is a significant concern about cancer recurrence after localised prostate cancer treatment [26]. Of all patients treated with direct prostatectomy, nearly 35% are reported to experience a relapse of PSA serum, after its total removal. This state is defined

as BioChemical Recurrence (BCR) and consists of new levels of PSA serum, probably due to residual prostate tissues and, most likely, recurrent prostate cancer [27; 28].

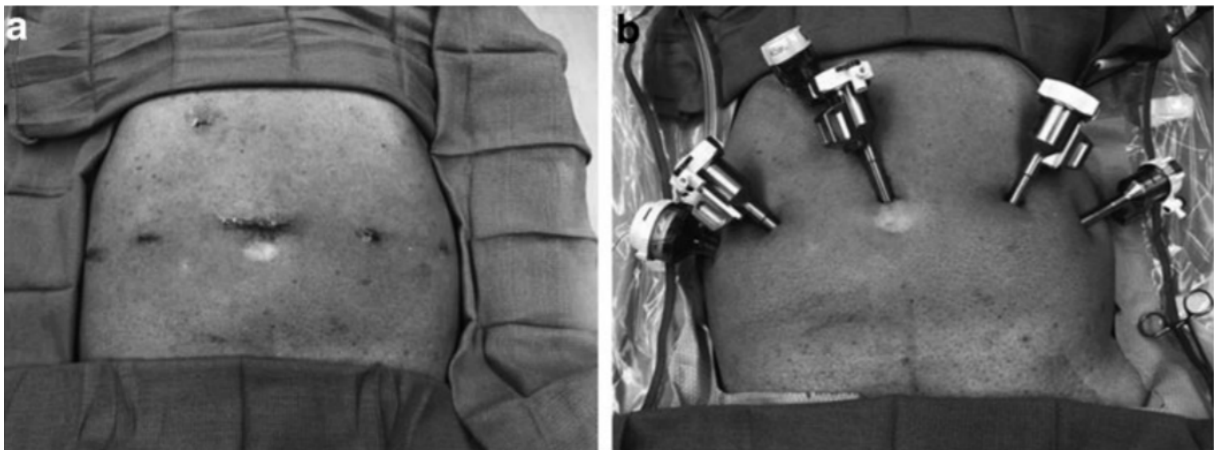

Figure 2.3: RALP. **a** Patient positioning and **b** Port Placement [25].

The emergence of BCR after radical prostatectomy has several causes and is highly variable [29]. Therefore, the clinician evaluation about PSA relapse is extremely important, as well as the distinction of the respective BCR based on preoperative or/and postoperative parameters. The decision will aid towards clinical decision-making, by introducing complementary therapies, achieving, thus, a more cost-effective and personalised treatment regimen.

2.1.5 Procedures Adopted in Hospital da Luz Lisboa

At Hospital da Luz Lisboa, mp-MRI is used to localise prostate lesions. The exam takes 30 minutes and includes three T2WI sequences with contrast for each of the three planes (sagittal, axial and coronal), two DWI sequences and a sequence of T1 Weighted-Image (T1WI) to check for possible presence of adenopathy.

Each exam takes approximately 45 minutes to be analysed in order to detect and localise the most relevant prostate lesions. Index lesions are defined as the lesions with the highest cancer suspicion score based on initial mp-MRI, irrespective of size. These lesions have an impact on patients outcomes by dictating the stage of the cancer. The stage can be defined as T1 (tumour too small to be seen on a scan), T2 (tumour found only in the prostate), T3 (tumour has spread outside the prostate on one side and into the tissue just outside the prostate) or T4 (tumour has spread into nearby structures other than the seminal vesicles). The stage classification is relevant because it is through this evaluation that the surgeon decides the surgical procedure (e.g. prostatectomy).

Conversely, identifying non-index lesions is complex and a significantly time consuming task for professionals. However, if a non-index lesion is near the border of the prostate, the stage of the patient can be reconsidered and sometimes changed from a T2 to a T3 or T4. This is relevant because, at a T3 or T4 stage, both prostate and margins should be removed in order to avoid a relapse of the disease.

Moreover, to classify and confirm the lesion, it is always necessary to perform a biopsy, an invasive

procedure.

When all the information about the biopsies and MRI analysis is available, then the surgeon is capable of planning the procedure, for example, whether it should be prostatectomy or radiotherapy to eradicate cancerous cells. After treatment, there are always post-surgery evaluation appointments, where the patient is followed in order to prevent any possible relapse.

2.2 Deep Learning

This section is mainly focused on depicting deep learning essential concepts. Firstly, basic concepts of neural networks, such as the perceptron and a more evolved network, the multi-layer perceptron, are described. The several processes of neural network training are also discussed in greater detail, as well as the most common evaluation metrics of generic networks. This sub-chapter ends with the description of CNNs, a specific type of neural networks which has been successfully implemented in different tasks, particularly those involving images.

2.2.1 Basic Concepts of Neural Networks

In the past few years, AI has been a subject of significant media hype. Machine learning, deep learning and AI usually come up in numerous articles with the promise of a future of intelligent chatbots, self-driving cars and precision medicine. Since its inception in the 50s, AI has developed and proven capable of addressing a wide variety of problems, and there remains considerable scope for further progress [30]. In particular, deep learning offers a different perspective on feature learning and representations, where abstract, robust and invariant features are hierarchically considered, from raw data. Despite being a subfield of machine learning, the concrete difference is how features are extracted. Traditional machine learning approaches, generally, need handcrafted engineering of features, while deep learning solutions have the capacity to automatically learn the best possible set of features intrinsically. Deep learning algorithms are based on neural networks, which in turn are inspired by biological neurons. Neural networks are then a set of layers hierarchically structured and connected in a chain. This area of research started with the perceptron network, which only has one layer, but over the years there has been a considerable improvement in the area and it is now possible to create deeper (i.e. with more layers) neural networks.

2.2.1.1 The Perceptron

According to Frank Rosenblatt, pioneering in the field of AI, the perceptron is a learning algorithm with a single layer [31]. As shown in Figure 2.4, a perceptron contains inputs, weights, bias, activation function and an output. The weights have specific values and a positive weight represents an excitatory connection, while a negative weight represents an inhibitory connection. By varying the weights, the function which the perceptron is able to compute is changed. The learning process relies on increasing or decreasing the weights of the active inputs when the output of the perceptron does not correspond to

the expected value. The bias is added and allows adjusting the point at which the artificial neuron will produce an output.

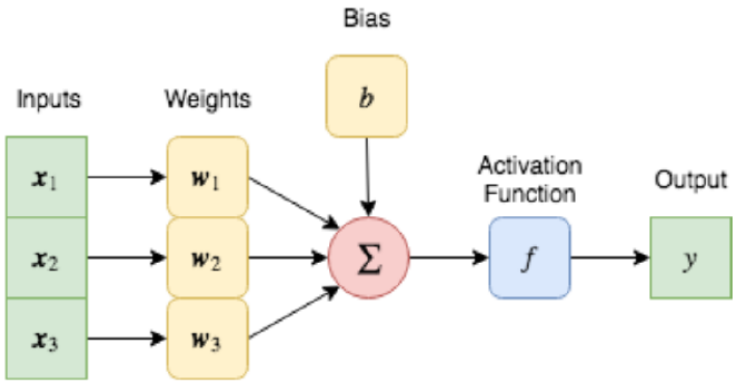

Figure 2.4: Perceptron model [32].

Mathematically, the perceptron model can be described as:

$$y = f(x) = g\left(\sum_{i=1}^{n} x_i w_i + b\right) = g(x \cdot W + b) \tag{2.1}$$

where *y* denotes the output prediction, *x* refers to the vector of input features, *w* is the vector of weights, *b* is the bias, and *g(.)* is the activation function.

2.2.1.2 The Multi-Layer Network

Neural networks are sophisticated computing systems that are able to perform complex tasks. While the perceptron only has one layer, the idea of a neural network consists of increasing the number of layers and also the number of nodes in each layer. This extension is the process that allows the completion of complex tasks, such as computer vision, speech recognition or machine translation. Figure 2.5 illustrates an example of a neural network with an input layer, hidden layers and a final output layer. The first layer is responsible for holding values of the input vector, while the latter holds the final results provided by the respective model. The hidden layers, the ones between the first and last layers, do not have specific behaviour and values, i.e. the learning algorithm is responsible for deciding how these layers should be used and implemented in order to get a considerable approximation to the desired output. The depth of the model is determined by the overall number of hidden layers in the architecture. Mathematically, a two-layer network can be described as:

$$y = f(x) = g(g'(x \cdot A + a) \cdot B + b) \tag{2.2}$$

where *A* and *B* represent, for example, the weights of the first and second layers, *a* and *b* the respective biases and *g(.)* and *g'(.)* the proper activation functions.

The fundamental component of neural networks is the artificial neuron, which can be associated with the perceptron. In general, a common neural network has several layers (depth) and several nodes (width) in each layer. The behaviour of each individual node is similar to the perceptron model, since it receives a set of values (input), it has to arrange the optimum set of weights and bias, and it also has an activation function, which determines if it produces an output.

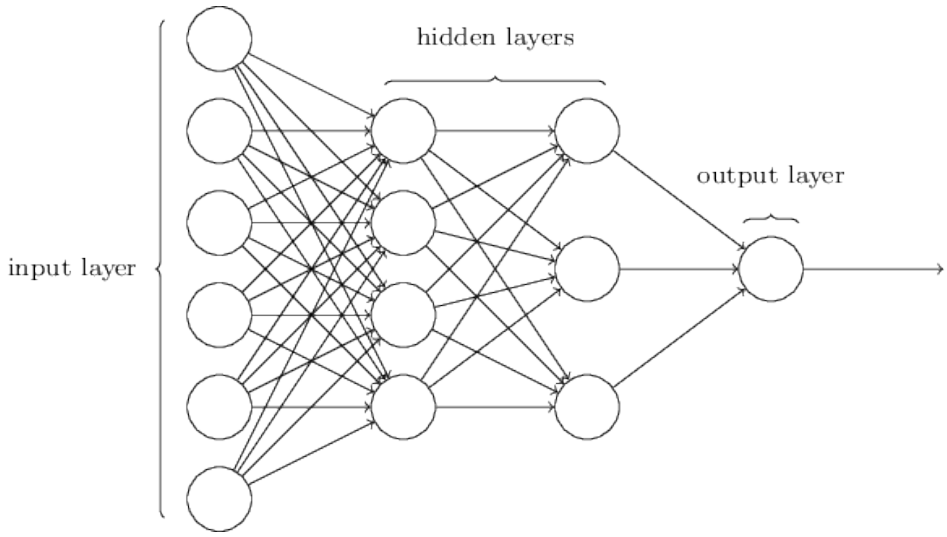

Figure 2.5: Schematic of a Neural Network [33].

Activation functions are important to introduce non-linearity properties to artificial neural networks. Without them, it would not be possible to perform complex tasks. Artificial nodes calculate a weighted sum considering their input, add a bias and then apply the activation function which decides whether that specific node should be activated and considered by the nodes of the next layer. The most frequently used activation function is a sigmoid. One advantage of this function is the fact that it produces a value in the range of 0 to 1 when an input from $-\infty$ to $+\infty$ is applied. However, the sigmoid function has the problem of the vanishing gradient, where the derivative values of the initial layers are very small and converge to zero during the training phase. Then, the learning process becomes very slow and the optimisation algorithm that minimises the error can reach a local minimum, not fully-exploiting the maximum performance of the neural network. Due to this issue, other activation functions can be considered, such as the Rectified Linear Unit (ReLU), in which the calculation is less computationally intensive when compared to the sigmoid function [34].

2.2.2 Neural Networks Training

A critical part of deep learning is the learning process, where the network is trained to reach the optimised values (weights and biases) to solve a certain problem. Back-propagation is a gradient-based algorithm which is used to optimise the parameters of the model, including its weights and biases [35].

The first stage of neural network training is the forward-propagation. This process occurs when the training data crosses the entire architecture, from the input layer to the output layer, in order to calculate the predictions (labels). At this stage, the input data passes through the network and each neuron applies a

specific transformation to the information received from the previous layer and sends it to the next layer. At the end of the network, after the information has passed all layers, the result is calculated in the output layer and a label prediction is created for the original input samples.

In the meantime, a loss or cost function should be selected to evaluate the performance of the model. It compares the output of the model, the predicted value, with the real value during the training process, in order to estimate how successfully the network is predicting. There are several standard loss functions that are used for this purpose, such as the mean square error, however, these functions can also be created by considering the metrics applied to evaluate the model. Once the loss has been computed, this information is propagated backwards (back-propagation) in order to improve the model predictions. From the output layer to the input layer, the gradient of the loss function is measured with respect to the weights and biases of each neuron. Therefore, these parameters should be updated in the opposite direction of the loss function gradient and optimisation functions are extremely important to indicate how much the values should be shifted. The most common optimisation function is the gradient descent, where the weighted matrix is initialised randomly and applied to the cost function. Parameter update through gradient descent can be mathematically described as:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_\theta J(\theta) \tag{2.3}$$

where θ represents the group of parameters, η the Learning Rate (LR) and J the loss function.

One of the most popular and powerful optimisation algorithms is the Adaptive Moment (AdaM). AdaM stands for adaptive momentum and it combines the momentum process with the root-mean-square propagation. With this approach, it is possible to reach the maximum performance of our model faster, without having the gradient descent issues, by calculating adaptive LRs for each parameter [36]. This optimisation function is shown in the following equation:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t} + \epsilon} m_t \tag{2.4}$$

AdaM calculates an exponentially weighted average of past gradients and stores them in variables m_t (with bias correction), as well as, an exponentially weighted average of the squares of the past gradients and stores it in variables v_t (with bias correction). Parameters optimisation is then updated in a direction based on the combined information from the previous gradients. In Equation 2.4, value ϵ is a small number simply to avoid division by zero. The respective gradients can be calculated by the following equations:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{2.5}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{2.6}$$

where β_1 and β_2 are hyperparameters that control the two exponentially weighted averages, g_t is the gradient on the current batch, while v_{t-1} and m_{t-1} are the variables before bias correction.

The described process is then repeated for each example of the input data and the weights and biases values are gradually updated until good predictions are obtained. Ideally, the process is repeated until the loss function is as close to zero as possible.

2.2.2.1 Hyperparameters Tuning, Overfitting and Regularisation

In the practice of deep learning there are model parameters and model hyperparameters. Model parameters are the properties of the model which are adjusted during training. Conversely, model hyperparameters are the properties which govern and structure the network training (e.g. LR and the number of hidden units) and their optimisation is a crucial task. The goal of the training process is to achieve a point where the results are good not only for the training set, but also for a validation set [37]. If the network is trained for too long, it is likely to become overfitted to its training set, ultimately underperforming when validated, just as the following figure indicates:

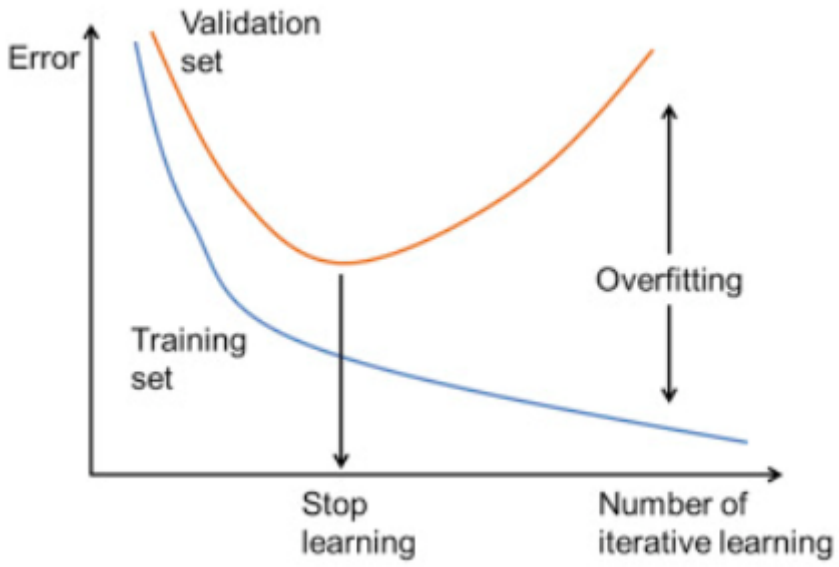

Figure 2.6: Effects of an excessive training period [32].

As shown in Figure 2.6, at a certain point (the "stop learning" point), if the number of epochs (i.e. the number of times that the model passes through the entire dataset), keeps increasing, the validation set error starts to increase while the training error keeps decreasing. After this point, the network is considered overfitted.

In order to avoid overfitting, there are several regularisation techniques. One of them is the dropout. It consists of ignoring a certain set of units (nodes), chosen randomly, during the training stage. By ignoring them, it means that these units are not considered during a specific forward and backward pass. In this way, nodes do not develop co-dependency with each other, leading them to achieve their maximum individual power and less overfitting, which means an overall smaller validation error [38].

Another technique to reduce overfitting is to add more data to the training set. One method to increase the amount of usable data is data augmentation. This is a simple task where the original images are transformed by varying features, such as: size, spatial orientation, mirroring, elastic transformations,

colouring, etc. As a result, the size of the training set is increased and it becomes more difficult for the model to overfit [39].

Underfitting is also a problem when training a network. A common solution to this problem is to increase the depth of the network (i.e. adding more layers). As a result, the nodes become capable of achieving their maximum potential and reaching the optimal training point [39]. However, increasing the network depth too much can also enhance other issues. When a network has too many layers, it becomes harder to tune all the parameters if the dataset is small. Additionally, as a model becomes excessively large, the training process can become extremely costly in terms of computing and storage, making the training period unreasonable.

Another hyperparameter that can be optimised is the LR, which depends on two aspects: batch size and the number of epochs. The batch size refers to the number of samples that are propagated through the network before each update of the network's parameters. Conversely, the number of epochs corresponds to the number of times that all training data passes through the network. These two influence the LR, which is a hyperparameter that controls how much the model changes in response to the estimated error each time the network weights are updated. If the LR is too low, then the model does not show any signs of progress in a reasonable time, whereas if the LR is too high, the model has difficulty in converging to the optimal point. A common solution is to update the LR hyperparameter during the training process. Also, to avoid excessive lengthy training, an early stopping strategy is usually implemented. This is a simple function that stops the model training if the network does not show any improvements over recent epochs. The number of tolerated epochs without any improvement is called patience [40].

2.2.3 Evaluation Metrics

Choosing the right metric is crucial while developing and evaluating deep learning models. Just as the loss function, the evaluation metric generally depends on the type of project under consideration. Regarding classification tasks, accuracy is the most used metric to evaluate model performance, which is simply defined as the number of correct predictions divided by the total number of predictions. This metric can also be more precise if a confusion matrix is created, allowing the user to check the model accuracy for each class. Recall is another important metric for classification models described as the fraction of samples from a single class that is properly predicted by the model. The Receiver Operating Characteristic (ROC) curve, as well as the Area Under the Curve (AUC) are two measures of performance of a binary classifier. AUC calculates the area under the ROC curve and is considered the probability of the model to rank a random positive example more highly than a random negative example. Conversely, ROC essentially represents the true positive rate against the false positive rate for different thresholds.

Deep neural networks can also be applied to regression problems. Regression models are generally developed to predict continuous target values, such as house price prediction. In these cases, selecting the proper evaluating metric is also extremely relevant. Mean Square Error (MSE) is probably the most popular metric used for this type of models and essentially finds the average squared error between the predicted and real values. Mean Absolute Error (MAE) is another important metric and is used to find the average absolute distance between the predicted and target values.

The most popular metric used for evaluating segmentation networks is the Dice Similarity Coefficient

(DSC) [41]. The DSC measures the overlap of two subsets, where 1 is the resulting value if the subsets are totally overlapped and 0 if there is no overlapping. Considering X and Y as the cardinalities of two distinct subsets, the DRE equation can be mathematically defined as:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{2.7}$$

Another metric that is used in the segmentation field is the Hausdorff Distance (HD) [42]. This metric measures how far two subsets are from each other in space. Only the highest value of all the distances, from a point in one set to the closest point in the other set, is considered.

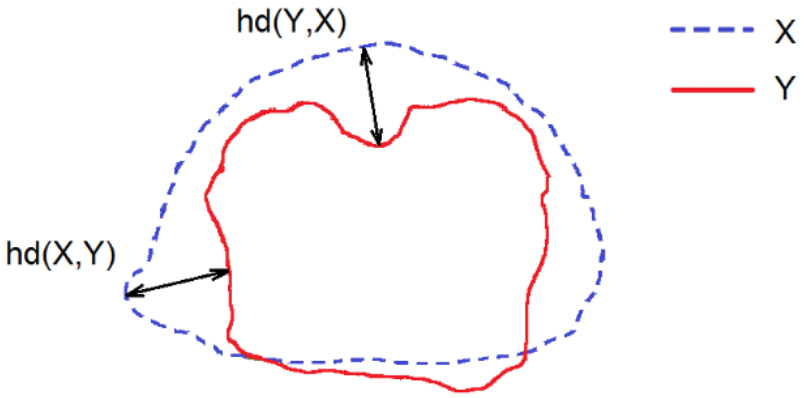

Figure 2.7: Hausdorff distance of two subsets [43].

The third most used metric in segmentation is the Mean Square Distance (MSD) [41]. It is similar to the Hausdorff metric, but it considers all the distances from a point in one set to the closest point in the other set, producing, in the end, their mean.

Although not applied in this dissertation, there is also the Intersection over Union (IoU), a metric that corresponds to the overlapping area divided by the union area of two distinct subsets [44].

2.2.4 Convolutional Neural Networks

Deep learning methods, specifically CNNs, have been successfully applied in the field of medical imaging to segment the anatomy of interest [45].

A CNN is a type of artificial neural network that is used to solve problems involving image recognition, object detection and other computer vision applications. Instead of using the common multiplication of matrices, a convolution - a type of linear operation - is completed in at least one of its layers. Equation 2.8 describes mathematically the convolution process.

$$s(t) = (x * w)(t) = \int_{-\infty}^{+\infty} x(\tau)w(t-\tau)d\tau \tag{2.8}$$

In Equation 2.8, the convolution is presented as *, where the first element, x, is the input and the second argument, w, is the filter. The output is commonly denoted as the feature map. The two dimensional convolution is represented in Figure 2.8.

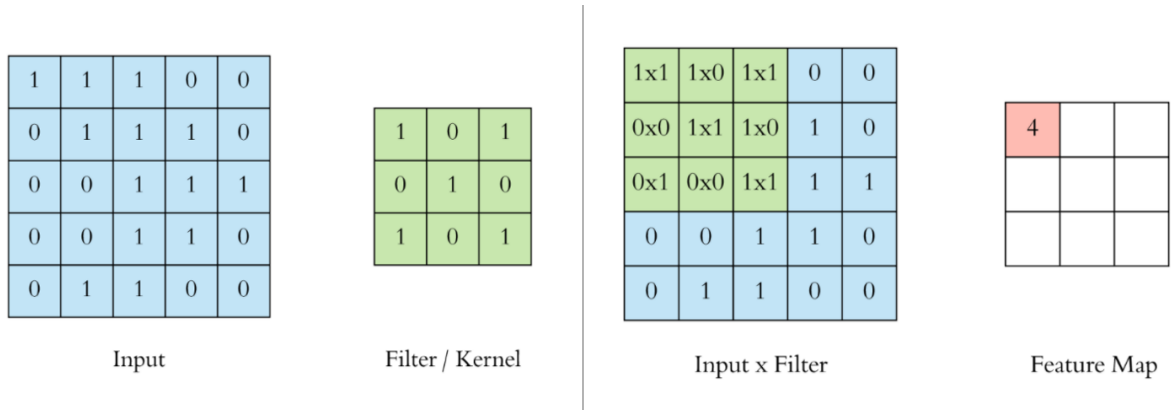

Figure 2.8: Schematic of convolutional operation with one filter of size 3x3, without padding and stride = 1.

Conventional neural networks usually take as input a *(n x 1)* vector, in which *n* is one of the dimensional parameters of the vector. However, the inputs of CNNs are vectors of *(n x m x 1)* for grayscale images or *(n x m x 3)* for coloured images, where the number 3 represents the RGBs (red, green and blue) components of each pixel. After the input layer, a CNN also has convolutional layers, where convolutional filters are applied to the input of that layer. Then, to reduce spatial dimensions, there are pooling layers. The most common pooling process is the max-pooling, where the highest value of a certain section is kept. In the end, there is a fully connected layer, where the output is adjusted by creating an *n*-dimensional vector, in which *n* is the number of classes considered in the problem. If the result should be a segmented image, at the end of the network, an output with the same size as the input is found. Figure 2.9 represents the flow of a CNN architecture with the example of a network which classifies images of digits.

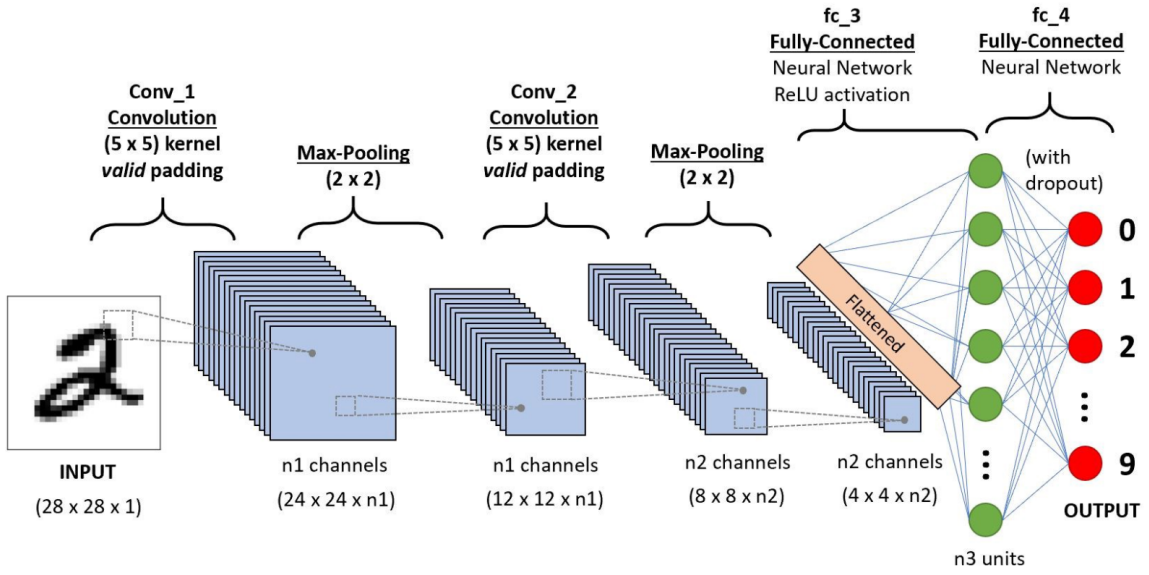

Figure 2.9: CNN architecture [46].

The convolutional layers and their respective filters have properties that are relevant to be enhanced during the convolution procedure:

- **Padding** is the process where extra pixels are added to the image (the most common are zeros - zero padding) in order to avoid the shrinking of the output every time a filter is introduced [47]. The dimension of the output is then:

$$n + 2p - f + 1 \tag{2.9}$$

where n is the dimension of the input, p is the number of pixels introduced and f is the dimension of the applied filter.

- **Stride** is a parameter that defines how far the filter moves in each step, vertically or horizontally. It is also important to adjust the dimension of the output or to minimise time during convolution processes. As a result, the output dimension considering stride and padding should be given by:

$$\frac{n + 2p - f}{s} - 1 \tag{2.10}$$

where s is the number of pixels that are shifted during each convolution [48]. This process is shown in Figure 2.10.

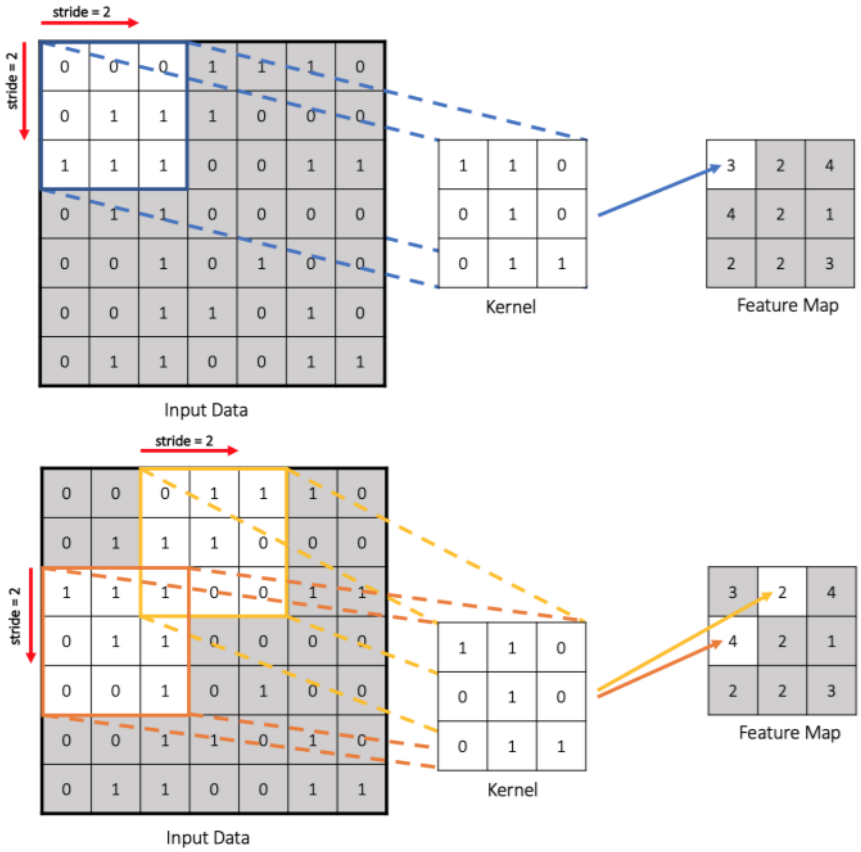

Figure 2.10: Schematic of a convolution process with a kernel of 3 × 3 and a stride of 2. The kernel slides through the input and at each position an element-wise multiplication and addition is performed, originating the feature map of size 3 × 3.

3. *Deep Learning for Image Segmentation*

Chapter 3 aims to describe the impact of deep learning for image segmentation by reporting the evolution of CNNs and its direct influence in medical imaging. Firstly, a description of deep neural networks for segmentation is presented, followed by its development and implementation in medical imaging with the U-Net architecture and respective variations.

3.1 Deep Neural Networks for Segmentation

Deep convolutional networks were first implemented by Y. LeCun et al [49] to recognise handwritten zip codes, in 1989. This was definitely a breakthrough, however, their success was limited due to the size of the available training sets, the small size of the considered networks and the limited computational power available at the time. In the last decade, through the advances in technology, deep convolutional networks have outperformed the state-of-the-art in many visual recognition tasks. It all started with Krizhevsky [50]. He and his team developed the first supervised training network with 8 layers and millions of parameters on the ImageNet database with more than 1 million training images. This was the first time that a CNN achieved the best performance for image classification in the ImageNet challenge.

The typical application of CNN are classification tasks, where the input is an image and the output is a single class label. Nevertheless, in many visual tasks, especially in biomedical image processing, the desired output should include the contours of the targeted zone, i.e. a class label needs to be assigned to each pixel. As a result, Ciresan and his group decided to train a network in a sliding-window setup with the purpose of predicting the class label of each pixel by providing a local region (patch) around that pixel as input [51]. This network was developed to segment neuronal membranes in electron microscopy images and won the ISBI 2012, an international EM segmentation challenge, by a large margin.

Yet, the Ciresan strategy has two drawbacks. Firstly, the network is very slow because it has to run each patch independently, creating a lot of redundancy due to overlapping patches. Secondly, there is a trade-off between the size of the image and location accuracy. Larger patches require more max-pooling layers, reducing the location accuracy, while small patches show better results in terms of location accuracy. Recent approaches have reached better results by proposing a classifier output that considers features from multiple layers: a block [52; 53]. Hence, it was possible to reach good location accuracy in images with several structures.

In 2015, the Computer Science Department of University of Freiburg, Germany, [54] created the U-Net architecture, the most popular network for segmentation. This model was developed to compete and win the ISBI cell tracking challenge in 2015 and still has, to this day, several applications.

The dataset used to test this network comprised the images available for the ISBI 2015 challenge, which consisted of 35 and 20 partially annotated training images of Glioblastoma-astrocytoma U373 cells and HeLa cells on flat glass, respectively. IoU was the metric used to evaluate this challenge and the U-Net has shown excellent results when compared to the other optimised networks, as Table 3.1 indicates. The U-Net performance on different biomedical segmentation applications was very good and researchers started to use it as standard architecture for networks used for segmentation [54].

Table 3.1: Segmentation results of the ISBI cell tracking challenge 2015 in terms of IoU [54].

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

The idea behind the developed architecture comes from the fact that in segmentation, it is not only necessary to convert feature maps into a vector, but also reconstruct an image from this vector. In this way, the same feature mapping can then be used to convert the vector back to an image. The feature maps that are developed during the contraction procedure are, respectively, added in the expansion section, preserving the structural integrity of the image and reducing the distortion considerably. The U-Net architecture can be represented as follows:

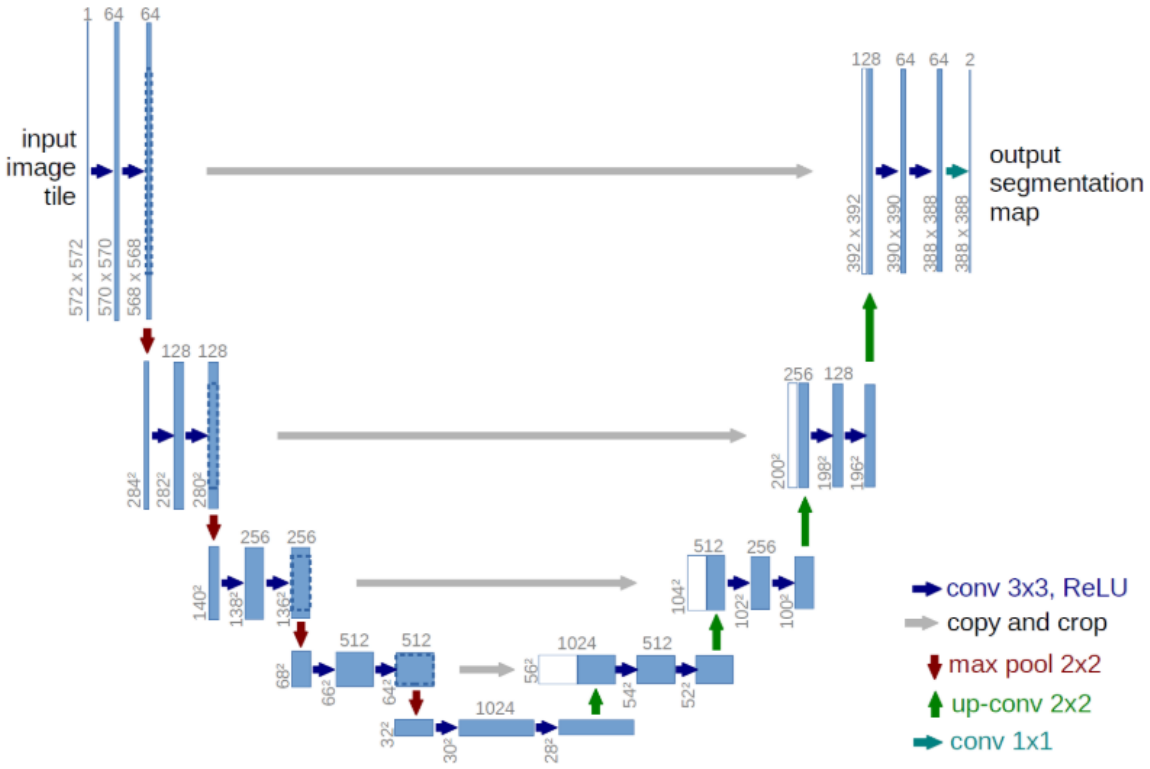

Figure 3.1: U-Net architecture [54].

The typical architecture for segmentation has three different sections: the contraction, the intermediate and the expansion section, as described in the following paragraphs.

The contraction section results from many contraction blocks. In this case, each block takes an image as an input, applies two [3 x 3] (unpadded) convolutional layers, followed by a ReLU and a [2 x 2] max-pooling operation with a stride value of 2 for contraction. The number of kernels of feature maps, after each block, doubles, and so, more complex structures can be learned.

The intermediate stage is set between the contraction section and the expansion section. It starts with the application of a max-pooling layer to the last result of the contraction section, then uses two [3 x 3] convolutional layers, followed by a [2 x 2] up convolutional layer.

However, the most important part lies in the expansion section, which also consists of several expansion blocks. Each block passes the input through two [3 x 3] CNN layers followed by a ReLU and a [2 x 2] up-sampling layer. Also, after each block, the number of feature maps used by convolutional layers is halved to maintain symmetry. In the expansion section, the output of each block is concatenated with the corresponding feature map from the contraction section. This action ensures that the features that are learned while contracting the image are then used to reconstruct it. The number of expansion blocks is the same as the number of contraction blocks. After that, a final [1 x 1] convolutional layer is applied to the resulting map to adjust each of the 64 components of the vector to the desired number of classes. In Figure 3.1, the represented network has 23 convolutional layers [54].

Meanwhile, several architectures were developed for segmentation on top of the U-Net. The most relevant segmentation architectures that followed U-Net are: 3D U-Net [44], V-Net [42], Adapted U-Net [55], Dense V-Net [41], HighRes3d Net [41] and Holistic Net [41].

In 2016, the same team that developed the U-Net, created an improved version of the network, the 3D U-Net. The 3D U-Net architecture is very similar to its 2D version, but instead of using 2D filters in the convolutional layers, it uses 3D convolutions. Another difference is the introduction of batch normalisation before each ReLU. Batch normalisation is a method to normalise the inputs of each layer, in order to fight the internal covariance shift problem. The model generated becomes more stable with this technique, allowing faster training processes and better performances [44].

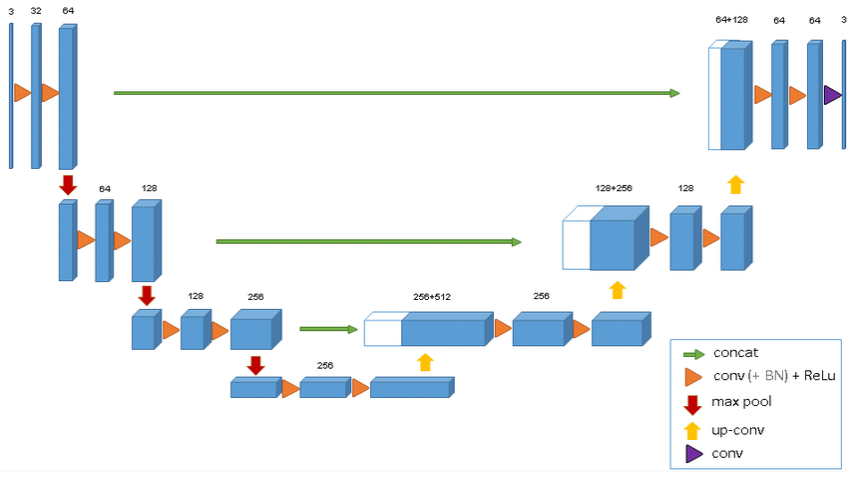

Figure 3.2: 3D U-Net architecture [44].

In the same year, 2016, a research team from the Technische Universitat Munchen, Germany, also proposed a 3D approach to solve problems involving 3D volumes: the V-Net [42]. The main difference between this architecture and the U-Net is the residual learning process. The implementation of residual learning consists of adding skip connections to jump over layers, accelerating the training process. This approach avoids the vanishing gradient problem that is responsible for slowing down the learning process by restricting the weights change during training. In Figure 3.3, the input of each block is used over the convolutional layers, processed through the non-linearities and added to the output of the last convolutional layer of that stage in order to enable learning of a residual function. This is relevant to save previous and valuable information during the contraction and expansion sections.

The V-Net model has the following architecture:

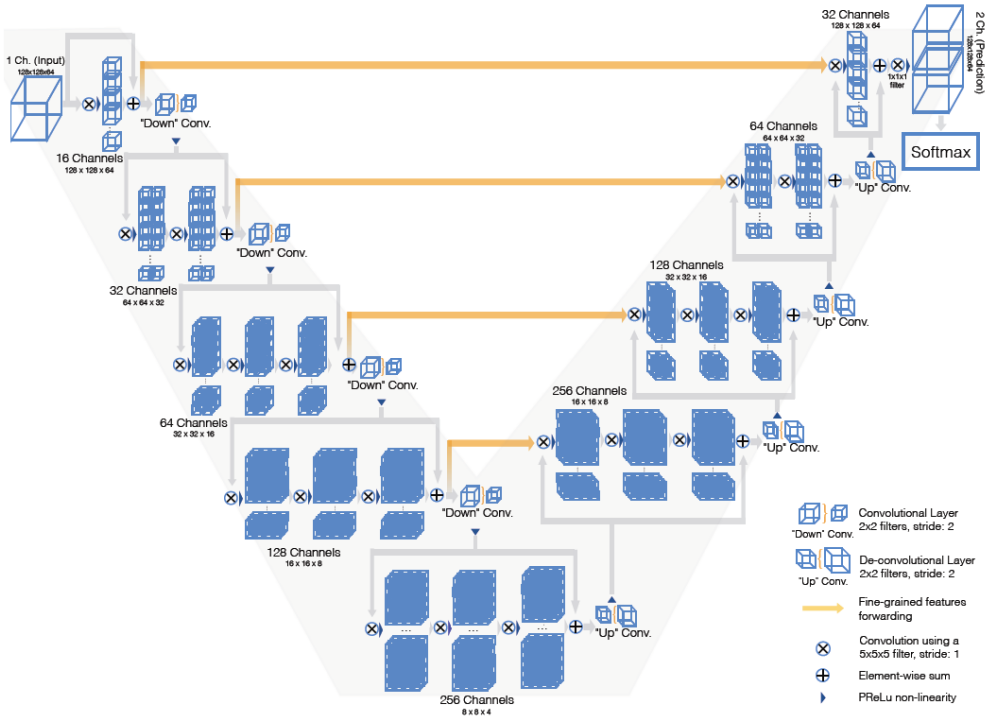

Figure 3.3: V-Net architecture [42].

In 2018, the Department of Medical Physics and Biomedical Engineering from the University College of London developed a more recent version of the U-Net: the Adapted U-Net architecture [55]. The Adapted U-Net has a similar configuration compared to the 3D U-Net, but with specific features according to the application field. The key of this architecture is the fact that before each down-sampling or up-sampling process, in each section of the network, a residual network (ResNet) block can be used, allowing the previous information to jump over some layers. This block does not change the number of channels nor the dimensions of the feature maps.

Several architectures continued to be developed over the years and shared with the research community. In 2019, the same team that developed the previous architecture wrote an article where the most relevant segmentation architectures are gathered and analysed [41]. The CNNs selected to be evaluated were: V-Net, Dense V-Net, HighRes3d Net, Holistic Net, Adapted U-Net and the original U-Net. The

difference, in terms of architecture, between the selected models can be seen in Figure 3.4:

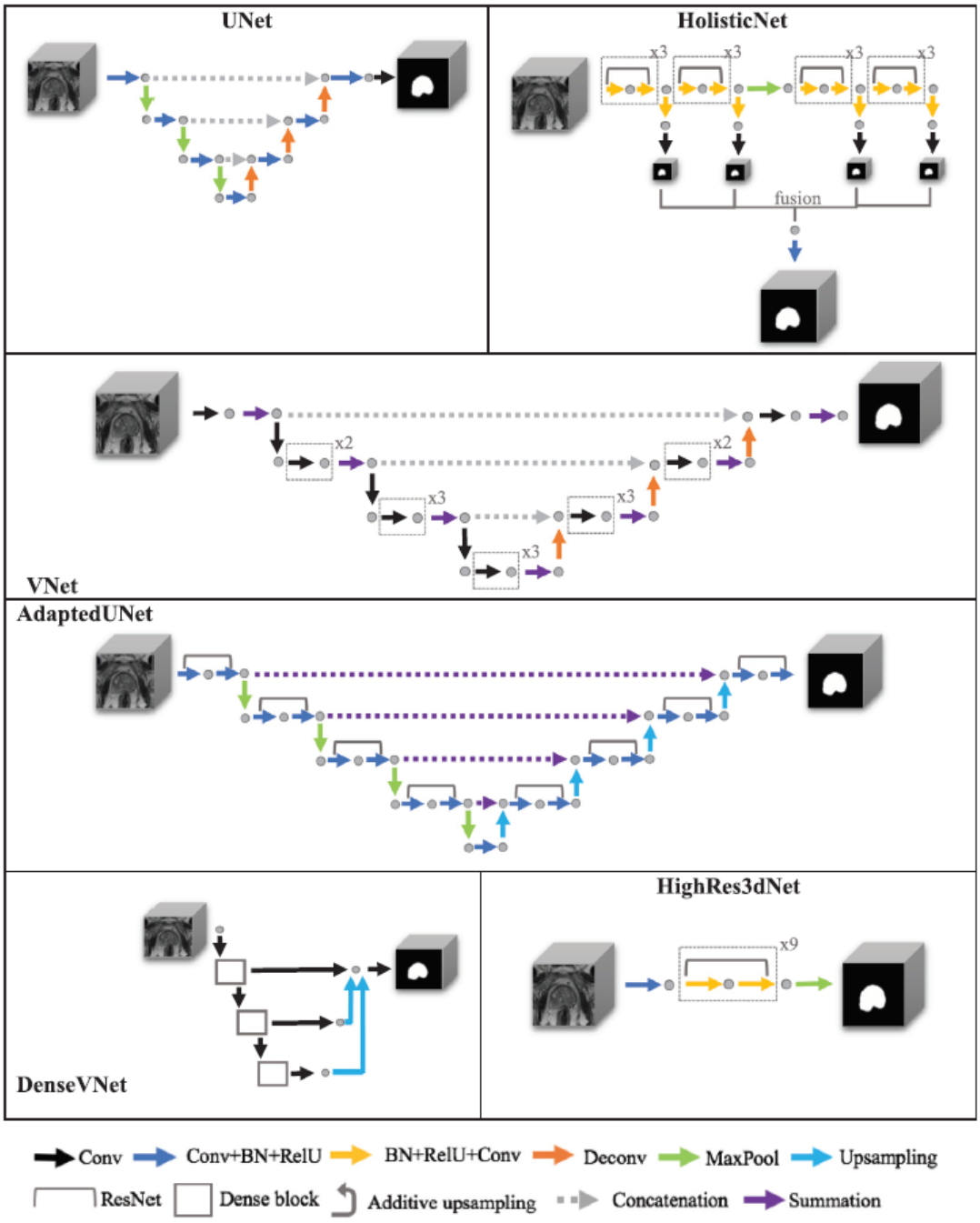

Figure 3.4: Architectures of the six models: U-Net, Holistic Net, V-Net, Adapted U-Net, Dense V-Net and High-Res3d Net [41].

As described in this section, nowadays there are several architectures for segmentation. U-Net has been the foundation of the following architectures, then more specialised and complex, allowing deep learning models to solve sophisticated problems.

3.2 Medical Imaging Application

In recent years, the use of CNNs in segmentation problems has become popular and their application to the medical field is an excellent example.

In 2016, the goal of the team who developed 3D U-Net created a CNN for volumetric segmentation that learns from sparsely annotated volumetric images. Whereas this approach is more difficult to implement and more costly in terms of computational power than its 2D version, it shows better performances for volumetric segmentation. The dataset used to train this network consists of three samples of Xenopus kidney embryos with 77 slices (2D) each. Similarly to the previous network, the IoU was the metric used as an accuracy measure to compare ground truth slices to the predicted 3D volume. The cross validation for fully-automated segmentation has shown an average IoU value of 0.723 when batch normalisation was not used in the network architecture and 0.704 when it was used [44]. These results are presented in Table 3.2:

Table 3.2: Cross validation results for fully-automated segmentation [44].

Test volume	3D w/o BN	3D with BN
1	0.655	0.761
2	0.734	0.798
3	0.779	0.554
Average	0.723	0.704

Conversely, the V-Net was proposed to solve clinical problems involving 3D volumes and participated in the PROMISE 12 challenge, a challenge where the goal was to segment MR images of the prostate [42].The CNN was trained end-to-end on MRI volumes depicting prostate and learned to predict segmentation for the whole volume at once.

Table 3.3: Quantitative comparison between the proposed approach and the 2016 best results in the PROMISE 2012 challenge dataset [42].

Algorithm	Avg. Dice	Avg. Hausdorff distance	Score on challenge task
V-Net + Dice-based loss	0.869 ± 0.033	5.71 ± 1.20 mm	82.39
V-Net + mult. logistic loss	0.739 ± 0.088	10.55 ± 5.38 mm	63.30
Imorphics [18]	0.879 ± 0.044	5.935 ± 2.14 mm	84.36
ScrAutoProstate	0.874 ± 0.036	5.58 ± 1.49 mm	83.49
SBIA	0.835 ± 0.055	7.73 ± 2.68 mm	78.33
Grislies	0.834 ± 0.082	7.90 ± 3.82 mm	77.55

In 2018, the Department of Medical Physics and Biomedical Engineering from the University College of London developed the Adapted U-Net architecture [55]. The idea was to create a deep learning method using CNN for automatic prostate segmentation in 2D Transrectal Ultrasound (TRUS) slices and 3D TRUS volumes. The dataset used consisted of TRUS images from 110 patients, where for each patient 38-177 para-sagittal slices were acquired (10-fold cross validation). The metrics used in this study were

the DSC and the MSD. The results yielded a DSC of 0.91 and a MSD of 1.23 mm. Although this approach was not applied to MR images, it was important to enhance the capabilities of the Adapted U-Net. In Figure 3.5 there are some examples of the Adapted U-Net application and a visual comparison with manual segmentation:

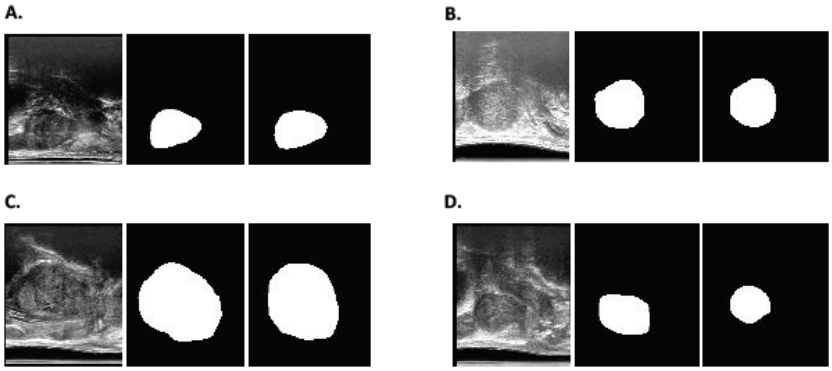

Figure 3.5: Examples of the manual and automatic prostate segmentation. The manual segmentations are shown in the middle and the automatic segmentations on the right-hand side. There are four different examples slices (A-D), chosen randomly and representative of all the images segmented [55].

In 2019, the same team that developed the previous architecture, wrote a scientific paper where the main goal was to compare the application of the most relevant models in segmenting the prostate in 3D patient T2-weighted MRI scans [41]. The CNNs selected to be evaluated, as referred to in the previous section, were: V-Net [42], Dense V-Net [41], HighRes3d Net [41], Holistic Net [41], Adapted U-Net [55] and the original U-Net [54].

In order to quantify the accuracy of the six architectures, a dataset of 232 patient magnetic resonance prostate images was chosen to be applied with a 5-fold cross validation. The metrics used to evaluate the performance of the networks were the DSC and the MSD. Different hyperparameter configurations (Value 1-Value 4) were used to tune the networks, as Table 3.4 shows, where the best configuration for each network was selected to be used for the final comparison [41].

After tuning each model, it was possible to compare the six CNNs. In terms of DSC, the HighRes3d Net yielded the best result with a value of 0.89 $\pm$0.03. However, the model with the lowest MSD value was the Adapted U-Net with 1.96 $\pm$0.61mm [41]. The results of each model are shown in Table 3.5.

Table 3.4: Hyperparameter configurations used for the tuning of the different networks [41].

Training Hyperparameter	Value. 1	Value. 2	Value. 3	Value. 4
Input image size	[112, 128, 64]	[80, 96, 48]	[48, 64, 32]	[32, 48, 16]
Initial learning rate	10^{-2}	10^{-3}	10^{-4}	10^{-5}
Weight decay	0	10^{-2}	10^{-4}	10^{-6}
Number of initial channels	4	8	16	32

Table 3.5: Segmentation performance of each network [41].

Network	3D DSC mean ± std [25th,50th,75th] percentiles	Boundary distance (mm) mean ± std [25th,50th,75th] percentiles
UNet	0.84 ± 0.07 [0.83,0.86,0.88]	2.52 ± 1.48 [1.73,2.07,2.57]
VNet	0.88 ± 0.03 [0.87,0.89,0.90]	2.45 ± 0.91 [1.78,2.36,2.88]
HighRes3dNet	0.89 ± 0.03 [0.88,0.89,0.91]	2.33 ± 0.81 [1.71,2.21,2.73]
HolisticNet	0.88 ± 0.12 [0.88,0.90,0.92]	2.56 ± 3.22 [1.62,2.04,2.50]
Dense VNet	0.88 ± 0.03 [0.86,0.88,0.90]	2.47 ± 0.66 [2.00,2.37,2.92]
Adapted UNet	0.87 ± 0.03 [0.85,0.88,0.90]	1.96 ± 0.61 [1.52,1.86,2.22]

In terms of imaging, T2WIs are the most used to detect prostate imaging features that indicate cancer [56]. T2WI provides useful delineation of the zonal anatomy of the prostate gland, although it lacks a high diagnostic accuracy when used individually. In this type of images, the peripheral zone of the prostate is shown as an area of high signal intensity whereas the central gland has variable signal intensity. Conversely, T1WI has poor differentiation between metastatic and normal tissue, being only important to detect whether there is internal bleeding near the prostate. Additionally, DWI is also used in prostate cancer diagnosis but as a complementary tool. Although DWI is not used by itself, it is an important supplementary sequence to assess the ease with which water molecules move around within prostatic interstitial space. This is relevant to detect neoplastic growths through changes of the regular water flow in certain zones. Since the objective of the dissertation is to delineate the prostate and its lesions, a combination of T2WI with DWI seems appropriate for this purpose.

At the moment, several networks are used for prostate segmentation. The most relevant models are adapted versions of the original U-Net. Some of these architectures, such as the original 2D U-Net, V-Net, 3D U-Net and Adapted U-Net are available on GitHub, a platform that provides hosting for software development. In this way, starting with the 2D U-Net architecture, and then develop/test other versions, seems the correct approach for prostate segmentation. Conversely, for prostate cancer segmentation with CNNs, the models should have the same structure as the models for the first task. This dissertation evaluates the impact of these neural networks not only to segment the prostate region, but also its lesions.

4. Materials and Methods

This chapter describes the materials and methods used to develop the CNNs used in the segmentation of the prostate region and then the segmentation of specific lesions. Firstly, a description of the data collection process is presented. Then, the conceptualisation of the models is detailed and discussed.

4.1 Data Collection

The data used in this project consists of prostate magnetic resonance images acquired at Hospital da Luz de Lisboa with a standardised protocol. Despite the standardisation of the protocol, in some situations to answer the clinical needs, some properties of the image, such as resolution, voxel space and number of slices, change according to the patient's size and examination requirements. Data acquisition consisted of T2WI and Apparent Diffusion Coefficient (ADC) series. All data was anonymised prior to analysis and informed consent was obtained. The dataset only included patients with biopsy confirmed prostate cancer. In this project, data from 53 patients were considered.

4.2 Data Annotation

In order to use these cases as masks to train and feed the networks, manual segmentation of each slice was previously completed by an experienced radiologist. The manual segmentations produced in each case were:

- Prostate segmentation in T2WI;
- Lesions segmentation in T2WI;
- Lesions segmentation in ADC.

The masks were produced in a prototype software from Siemens (Multi-parametric Analysis in Syngo Via), designed specifically for this purpose. Syngo Via is a multimodality reading solution built on a client-server platform which compiles several applications. The masks created with this software are binary images, with "ones" inside the region of interest and "zeros" elsewhere. The procedure followed this specific order:

1. Check the identification number of the patient and extract the corresponding magnetic resonance prostate exam in Syngo Via.

2. Select the T2 sequences, as well as the diffusion and ADC sequences, and open them with the Multi-parametric Analysis software, which is available in Frontier (UI where different medical applications can be accessed). The software automatically co-registers both images.

3. Create the masks of the prostate region. At first, the automatic prostate segmentation tool, provided by the software, is applied and then all slices are visually evaluated and edited if required. Small adjustments on the masks are generally necessary, especially in the apex zone.

4. Save prostate masks in Syngo Via.

5. Repeat the two previous steps (3 and 4) to segment prostate lesions in T2WI and ADC. The pathological report of each case is firstly consulted in order to localise properly the lesion in the image.

6. Segmentation data were saved

All the anonymised data was transferred locally, so the models could be trained in a personal computer.

4.3 Data Processing

4.3.1 Data Organisation

In order to process data, as well as to build all models, Python was the selected programming language while JupyerLab was the preferred environment to work with.

The data was stored slice by slice, per image, as Dicom files. Simple ITK and NumPy were the libraries used to read the Dicom files, convert them to arrays and then save them as *.mhd* volumes (Meta-Header format). At the end of this process, each patient folder had the following volumes with matching number of slices:

- Original T2WIs;
- Original ADC images;
- Prostate masks performed in T2WIs;
- Lesion masks performed in T2WIs;
- Lesion masks performed in ADC images.

In the dataset, most volumes presented the same bidimensional matrix size, [512 x 512], while the depth varies between 28 and 34 slices. Voxel spacing and offset values are also values that were set according to the patient characteristics.

4.3.2 Data Quality Evaluation - Segmentation

After reorganising the data, the images were ready to be processed. Prior to modelling, a data quality set was performed, where the data was carefully analysed. Small holes in the masks, probably due to human error during the manual segmentation or even prototype limitations, were noticed and needed to be filled. Some holes were supposed to be there, because they did not represent prostate or lesion, but others were misclassifications and needed to be filled. The solution was to apply a matrix of ones with a size of [3 x 3] across the image and the holes that were inside this matrix were filled. The library used to complete this process was Scipy.

4.3.3 Data Quality Evaluation - Intensity Ranges

The next procedure was an objective evaluation of the data. The ultimate aim of the project is to segment lesions in the prostate region, so the image intensity between the prostate and its lesions was analysed and compared.

The initial step of this task comprised reading the volumes of the original T2WIs, the prostate masks and lesions masks in T2. Then, the transformation of these volumes to arrays was completed in order to manage data more easily. A "for loop" was created to read volumes slice by slice and when the prostate was visible in the original image, the respective slices were stored as lists. In the end of the loop, three lists were created with: the original images, the prostate masks and the lesion masks for the slices where the prostate was present. After this procedure, multiplying the original images with the prostate and lesion mask was possible, yielding two volumes: one volume with the original images cropped around the prostate and another volume with the original images cropped according to the lesions localisation. The intensity pixels histograms were performed for the two volumes with the Matplotlib library. In order to evaluate histograms similarity, a spreadsheet with the difference between the intensity peaks in each case was created.

The next step of image processing was imaging normalisation in both T2 and ADC sequences. Normalising pixel intensity is important to create uniformity and consistency across the dataset. This process transforms the intensity pixel curve into a normal distribution. The Z-score normalisation was the procedure applied instead of the regular maximum and minimum normalisation to avoid the impact of possible outliers.

Mathematically, the Z-score normalisation can be described as:

$$Z = \frac{X - \mu}{\delta} \tag{4.1}$$

where X represents the original values, μ the mean value of pixels intensity and δ the standard deviation value.

4.3.4 Data Selection

The last procedure of the image processing was the data selection. In some cases, it was extremely hard to precisely detect the contours of lesions, since the intensity curve histogram of the prostate and lesion were almost identical in both T2WI and ADC. Each case was observed and validated and a sub-dataset was created with the most distinguishable cases. The criteria used were the curve shape of prostate intensity compared to the lesion curve, as well as the difference between the intensity peaks. If the curves were similar, and so were the peaks, the case was discarded.

In the end of this section, 28 cases were considered in the sub-dataset for the lesion segmentation task, with a total of 810 slices. Each slice could be visualised as original T2WI, original ADC image, prostate mask, lesion mask produced in T2WI and lesion mask produced in ADC image.

4.4 Model I - Prostate Segmentation

The overall objective of this dissertation is to automatically segment prostate lesions. MRI provides images with a broad vision over the prostate region and the lesions are usually just a small portion. This section describes a model for prostate region segmentation, in order to identify the contours of the prostate and crop the image. Consequently, the image becomes cropped and the dimension of the lesion is increased proportionally. This step might improve the training process of the lesion segmentation model by enlarging the region of interest.

According to the state-of-the-art, the selected reference architecture was the traditional U-Net [54]. Based on this architecture, a similar model was developed with an important addition: residual blocks. These blocks were already described in the concept description chapter (3) and they are particularly relevant to train a model with significant depth. The conceptualisation of the model was built in Python, specifically with Keras and Tensorflow as backend.

The architecture receives a slice at a time, so the input layer has the size: [512 x 512 x 1]. The initial weights and biases values are set randomly inside a certain range. It has four blocks of depth, which means there are four blocks in the down-sampling and four blocks in the up-sampling processes. Each block during the down-sampling begins with a convolutional layer with a kernel size of [3 x 3], padding type as “same” (i.e. adjustment of necessary padding values to have the output shape equal to the input) and a ReLU activation function in the end. This layer is followed by batch normalisation layer which applies a transformation that maintains the mean output close to 0 and the output standard deviation close to 1. The rest of the blocks is a repetition of these two previous layers. The blocks also have a skip connection between convolutions on each level, providing smoother loss curves and avoiding gradient-related problems. A concatenation layer is also present after each block, adding the initial information to its respective block in the up-sampling section. After each block of the down-sampling section, there is a max pooling layer that operates for 2D spatial data and down-samples the input shape by half of its dimensions.

After the four blocks of the down-sampling process, there is also a block with the convolutional and batch normalisation layers. However, instead of being followed by a max pooling layer, an up-sampling

2D layer is applied. This layer begins the up-sampling section of the network, which has the same number of blocks. Each block is designed to have the same number of convolutional and batch normalisation layers, with the objective of reducing the feature maps depth by half. After each of these blocks, an up-sampling layer is applied to increase to double the image size. Consequently, the output layer of the architecture has exactly the same shape as the input layer with the predicted segmentation.

With the architecture designed, there were other steps that needed to be followed. The first step was to manage the GPU options with the ConfigProto library. This was important to run the model with the local GPU, instead of the local CPU. Since neural networks training is primarily repeating matrix operations, training using GPUs is considerably faster.

The code describing the model was written in a JupyterLab notebook. This notebook presents four more sections:

- Augmentation;
- Metrics;
- Train;
- Test.

The augmentation section presents two functions related to the data augmentation method used during the training. The first function, built with the OpenCV library, induces distortion to the input images, creating new blurred slices with different shapes. With SimpleITK, the second function is responsible to denoise an image using curvature driven flow. By applying these two functions to the original input images, it is possible to create new and more slices with different shapes while maintaining the same original structure.

Conversely, the metrics section describes the several metrics used to evaluate the model. The main metric was the DSC, which measures the overlap of the predicted and original segmentations. The function created to measure DSC first flattens the input tensors (predicted and original segmentations) and then sums the multiplication of the flattened tensors to find the intersection value. There is also a sum of each flattened tensor to find the individual X and Y values of the equation, presented in Chapter 2. The other considered metrics were the HD and the MSD, both based on the surface distance function. This function also flattens the same input tensors and then measures exactly all distances between the original and predicted segmentation contours. The MSD is the mean of the resulting values and HD is the highest value.

Data preparation is an important initial step. The first step was the definition of the resize function to adjust the collected images shape according to the network input [512 x 512]. Some collected images came with sizes that were different from the expected [512 x 512] and with the OpenCV resize function, it was possible to adjust the size of those images using nearest neighbour interpolation. Pixel dimensions are also proportionally adjusted according to their respective resizes. The next step was the transformation of data to arrays. The network was designed to train [512 x 512 x n] images, where *n* is the number of all training data slices. In the designed function, four empty lists are created, representing the original T2WIs and respective prostate masks for training and for validation. Previously, the dataset was randomly split

80:20 in two folders, the first for training tasks and the second for validation. A "for loop" is used to read the folders where data was stored for training or validation and also to append each slice to the respective list. With the lists completed, they are concatenated and transformed to NumPy arrays with a [512 x 512 x n] format. In the end of this function, the four arrays are stored as *.npy* files.

The next step of the training section was the fit generator function definition. This function begins with the data loading, calling the four *.npy* files. Then, a dictionary is created to be used in the *ImageDataGenerator* function, selecting the transformations applied to the original images. The *ImageDataGenerator* is a Keras function that transforms original images each batch, resulting in real-time data augmentation. The considered dictionary leads the *ImageDataGenerator* to promote vertical and horizontal flips, height and width shift ranges, zooming level modifications, rotations and also the transformations presented in the augmenters section. The changes completed in the original images are equally completed in the respective masks. After selecting the previously described architecture features, model checkpoint and several callbacks needed to be completed. The model checkpoint defines where the model weights are going to be stored and which is the metric to optimise. In this case, the metric to monitor was the DSC loss of the validation set. The epoch with the lowest DSC loss of the validation set is stored. Callbacks define criteria where some model parameter are fine tuned. Two callbacks were created to increase the training process. The first type of callback refers that when the DSC loss of the validation set does not improve in 125 epochs, the training process is stopped as a result. This can be viewed as a type of regularisation, which can aid in the avoidance of overfitting. This callback is called *Early Stopping*. The second type of callback reduces the LR parameter value by a factor of 10 when the monitor metric has stopped improving for more than 80 epochs. This callback, the *Reduce LR On Plateau*, is important to reach minimum loss values, in terms of validation DSC. Once the callbacks are described, the model is compiled using the AdaM optimisation function and an initial LR of 0.01. The model fit generator is finally applied and the network is trained for 600 epochs.

In order to optimise the model, several networks were trained considering some changes in the architecture configuration and parameters values. The LR was tested between 0.1 and 0.001, batch size between 16 and 64, dropout between 0.5 and 0.8, architecture depth between 4 and 6 blocks and number of initial nodes between 2 and 6. Conversely, the steps per epoch value was fixed and obtained by doubling the result of the division of the number of images in the training set by the batch size. As a result, with *ImageDataGenerator*, the number of trained images is twice the original number (data augmentation).

The following section describes the test section. In this section, the results of the model, as well as the images with the predicted segmentations, are presented. There are functions to read data, to make plots, to check predictions results and also to load model weights.

The reading data function was created to read the original segmentation masks. These masks are then used to compare to the predicted segmentation masks, provided by the check predictions function. The data reading and preparation are similar to the function which converts data to arrays described in the section 4.4. The function to obtain the model was designed to load the stored weights of the trained model with the Keras function *model.load_weights*. The most important function of this section is the check predictions. This function starts by loading the four *.npy* files and the model weights. The files are converted to arrays. It uses the *model.predict* command on the original T2 validation images to

get the predicted prostate segmentations NumPy array. With the original and predicted arrays of the validation set, applying several metrics was possible. Each DSC, HD and MSD values are appended to a specific empty list and the lists are converted to arrays in the end. The mean DSC, HD and MSD of the validation set are obtained by applying the *np.mean* to the respective arrays. Besides the overall metrics, best predictions and worst predictions were evaluated empirically. In the end, the predicted prostate segmentations were stored as *.mhd* files in the "Predictions" folder.

4.5 Model II - Prostate Lesion Segmentation

This section presents the prostate lesion segmentation model. The idea behind this model is to receive cropped images according to the predicted prostate size determined by the previous model I, reducing the amount of non-prostate data presented to the networks.

The first task of this section was the crop function definition. This function reads the predicted prostate segmentations of the "Predictions" folder, converts them to arrays and registers the ideal crop region in each case. Original T2 and ADC images are loaded, as well as the predicted prostate mask and the original T2 and ADC lesion masks. With the predicted prostate masks, all the other loaded images are equally cropped. A "for loop" is created to analyse one slice at a time and the slices with the prostate mask identified (*np.max(mask_slice* == 1)), both X and Y maximum and minimum values are stored in empty lists. As a result, each list represents the maximum and the minimum X and Y values of each case. With these lists, determining the ideal crop coordinates is possible, using *np.max* and *np.min* functions. This process was completed with all the available dataset and a spreadsheet was created with all the X and Y crop ranges for each case - e.g. case of patient 02 with minimum prostate X value of 20 mm and maximum value of 120 mm has a range of 100 - plus a safety margin of 15 mm. The spreadsheet was analysed and, to create homogeneity, a single X and Y ranges were selected to be applied in the whole dataset. This selection was based on the highest range values for each coordinate. The selected ranges were actually the same to crop the entire dataset evenly: 256 mm in both X and Y coordinates. All the T2 and ADC original images and T2 and ADC lesion masks were then cropped, creating a new dataset with the prostate zone cropped and with a size of [256 x 256]. Another task to enhance the prostate region was removing the slices where the prostate was not detected. As a result, the number of slices was reduced, promoting a faster and concise training.

In this section, five models were trained to segment prostate lesions and ultimately compared. The considered models were different as they used different datasets, as follows:

- Model 1: T2 original images + T2 lesion masks;
- Model 2: ADC original images + ADC lesion masks;
- Model 3: T2 + ADC original images + T2 lesion masks;
- Model 4: T2 + ADC original images + ADC lesion masks.

These models datasets exclusively used cropped original images and masks. The first two models only received one input image to segment, so the architecture is the same as the model for prostate segmen-

tation. Conversely, the models with T2 + ADC original images as inputs, needed a network adjustment, since two images were being received at a time. Both ADC and T2 images were compatible and could be overlapped.

In terms of parameters definition and considering the best configuration of the prostate segmentation model, all models were set with equal values: dropout 50%, 600 epochs for training, 125 epochs for early stopping, 80 epochs for reducing the LR, initial LR of 0.01 and batch size of 32. The relevant difference was the architecture adjustment for the models which received two images at a time, as well as the respective dataset preparation. The "data to array" function was redesigned to create [256 x 256 x 2] arrays instead of creating [256 x 256 x 1] arrays. This type of array has two images, with the same 2D shape, concatenated and merged. The *.npy* files of the original images are stored with these type of arrays, while the masks remain independent, i.e. with a [256 x 256 x 1] shape. The other necessary change was the network architecture. With these *.npy* files as original images input, the architecture was adapted for the input layer, being capable to receive two images at a time. The output of the network was a single mask, so the information of both T2 and ADC original images is combined through the network.

In this final evaluation, data was not split randomly 80:20. Instead, K-fold cross-validation was applied with K = 4. Since Keras *fit_generator* - the function which fits the model - does not have a direct cross-validation implementation, the process was built manually. Data was split into 4-folds, each fold had seven cases. The function was executed inside a 'for cycle', loading the respective training and validation folds at a time. The DSC values are registered as *model_history* lists and each list is saved as a specific numpy array. In the end of the cross-validation process, eight numpy arrays were returned: four arrays describing the DSC evolution of the training set and four indicating the DSC values of the validation set. Each model presented four graphs with the DSC progress, representing each of the data combinations. After model training, mean DSC values were calculated based on the cross-validation DSC arrays. The mean DSC values were plotted and the highest DSC validation value is considered the mean DSC of the trained model. The best epoch of each K-fold combination was also registered and it was used to measure its HD and MSD values. The averages of HD and MSD values were considered the mean values of each fold regarding to these metrics.

In the end, four models were created and compared using the above mentioned metrics for lesion segmentation. Additionally, a single model with T2 original images and lesion masks also created with T2 images was built, however, without original images cropped. This is important to have as a reference and compare if lesion segmentation in two steps (firstly prostate segmentation model to crop the images and secondly the lesion segmentation model with zoomed images) is better than a one-step lesion segmentation procedure which does not include enlarging the prostate region.

5. *Results*

The major results are described in this chapter. Firstly, the data collected is presented and then the subsequent data processing and analysis are introduced. Several models were developed in this project with two different purposes: segment the prostate region and segment prostate lesions. The results of these models are listed in this chapter. As previously mentioned, all the data analysis and models training were performed using Python.

The first part of this dissertation was the data collection from Hospital da Luz Lisboa, followed by data processing. For each patient, the collected data comprised a volume of original T2WIs, a volume of original ADC images, a volumetric mask of the segmented prostate using T2WIs, a volumetric mask of segmented prostate lesions using T2WIs, and a volumetric mask of segmented prostate lesions using ADC images.

In Figure 5.1, an example of three slices of a patient's case are visualised in Python.

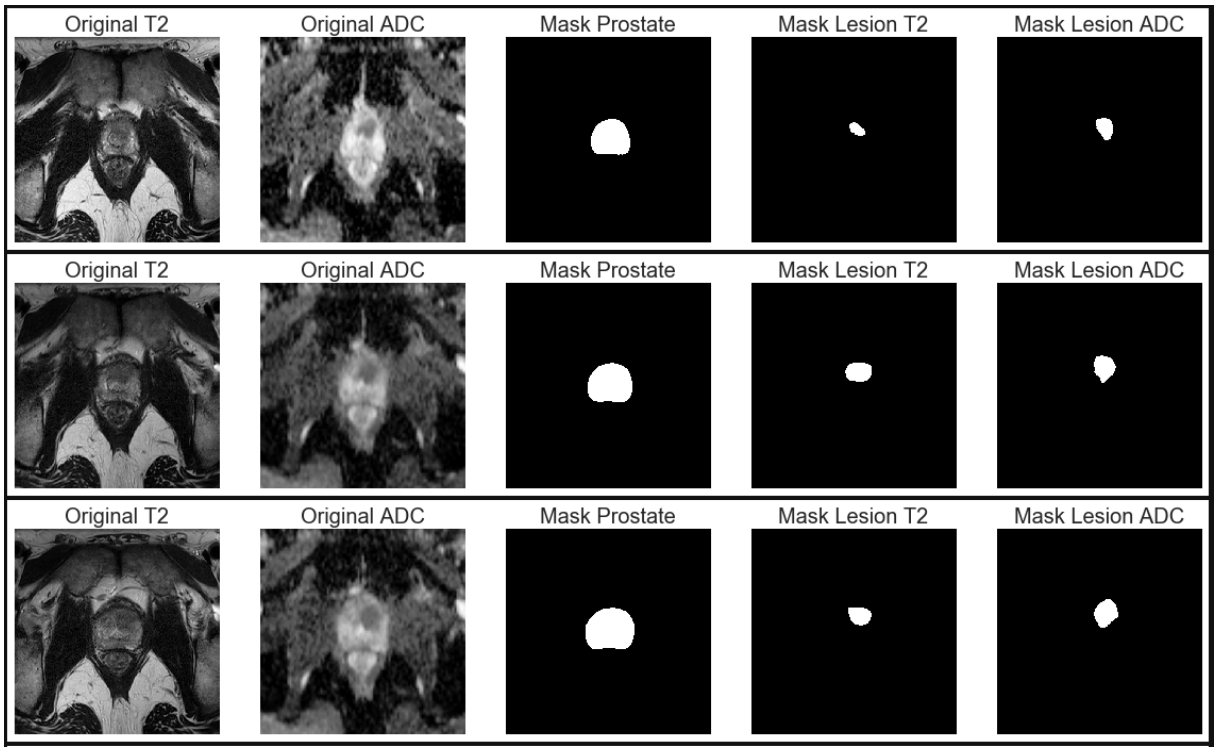

Figure 5.1: Example of the same three slices in T2 and ADC (original images) and their respective masks (prostate and lesions).

In Table 5.1, relevant information of ten patients' cases is described.

Table 5.1: Information of ten considered cases.

Name	Width	Height	Depth	Voxel Spacing
Case 01 Original T2	512	512	30	(0.27, 0.27, 3.00)
Case 02 Original T2	512	512	30	(0.27, 0.27, 3.00)
Case 03 Original T2	512	512	28	(0.27, 0.27, 3.00)
Case 04 Original T2	512	512	28	(0.27, 0.27, 3.00)
Case 05 Original T2	512	512	28	(0.27, 0.27, 3.00)
Case 06 Original T2	512	512	28	(0.27, 0.27, 3.00)
Case 07 Original T2	512	512	28	(0.27, 0.27, 3.00)
Case 08 Original T2	512	512	31	(0.27, 0.27, 3.00)
Case 09 Original T2	512	512	30	(0.27, 0.27, 3.00)
Case 10 Original T2	512	512	28	(0.27, 0.27, 3.00)

The next step was the process of filling holes in the the masks. Several masks of the prostate and lesions presented few holes clearly that were due to manual mistakes or software issues. In Figure 5.2 the results of this task are exhibited.

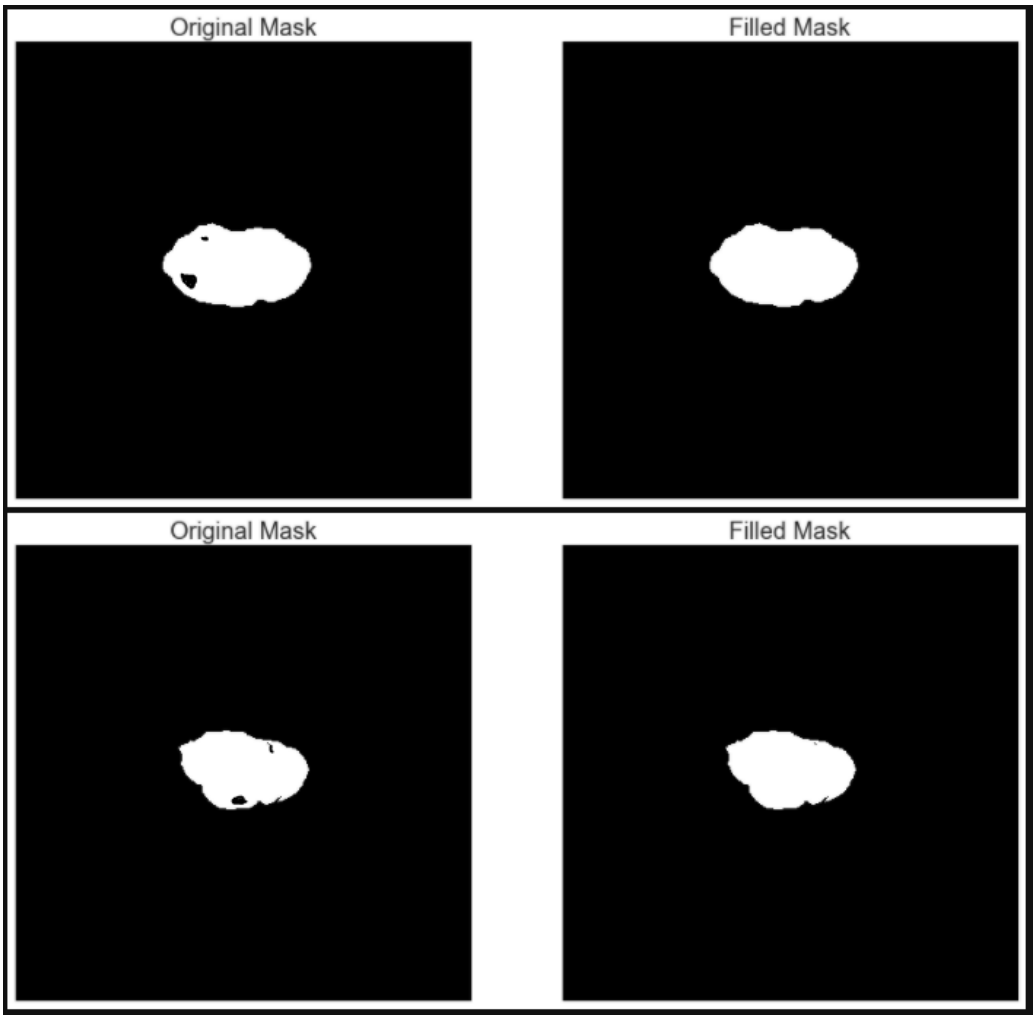

Figure 5.2: Differences between the original mask and the filled mask.

Analysing data is an important task in every project. Intensity pixels histograms were created in order to compare imaging details between the prostate and its lesions. This process was replicated across the entire dataset and the intensity peaks of both histograms were registered for each patient case. In Figure 5.3 both histograms of a patient's case are represented.

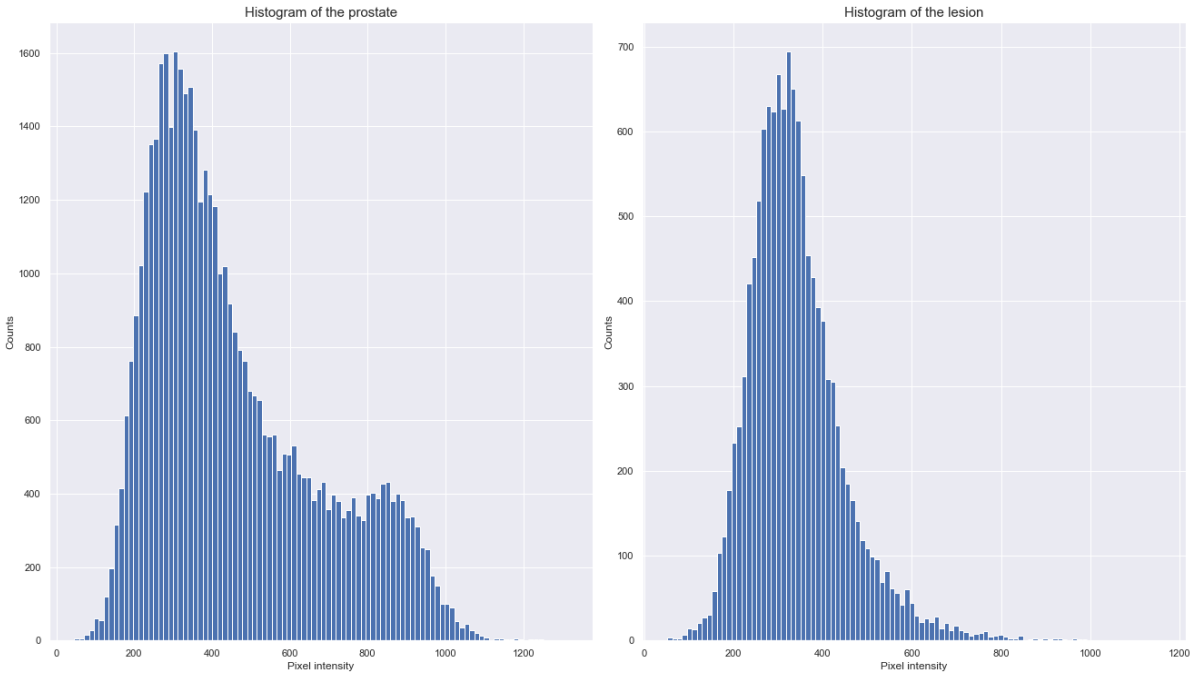

Figure 5.3: Histograms of the prostate and lesions intensity of patient IAP077.

According to the spreadsheet created to register the histograms peaks, the mean difference between those peaks when considering all dataset is only 67, the maximum difference is 243 and the lowest is 3.

After data normalisation, the last procedure of data processing was the data selection. As described in the previous chapter, some cases were not considered for prostate lesion segmentation. The dataset information after this selection is described in Table 5.2.

Table 5.2: Dataset information for each model target.

	Prostate Region Model	**Prostate Lesion Model**
Number of Cases	53	28
Number of Slices	1534	810

Once the dataset was processed, the first model was trained. Several parameters and architecture configurations were tested and compared, with the three most positive presented in Table 5.3. The best results for prostate segmentation are enhanced in Table 5.4, considering the DSC, HD and MSD state-of-the-art-metrics.

Table 5.3: Three best applied configurations for prostate region segmentation.

Prostate Segmentation			
	Trial 1	**Trial 2**	**Trial 3**
Initial Learning Rate	0.01	0.01	0.1
Dropout	0.5	0.5	0.5
Batch Size	32	16	32
Architecture Depth	4 blocks	5 blocks	4 blocks

Table 5.4: Results of the three best trials for prostate segmentation (validation set).

Prostate Segmentation			
	Trial 1	**Trial 2**	**Trial 3**
DSC	0.88	0.85	0.82
HD	16.5	18.2	17.7
MSD	2.1	2.5	2.4

The Figure 5.4 represents the training process of Trial 1.

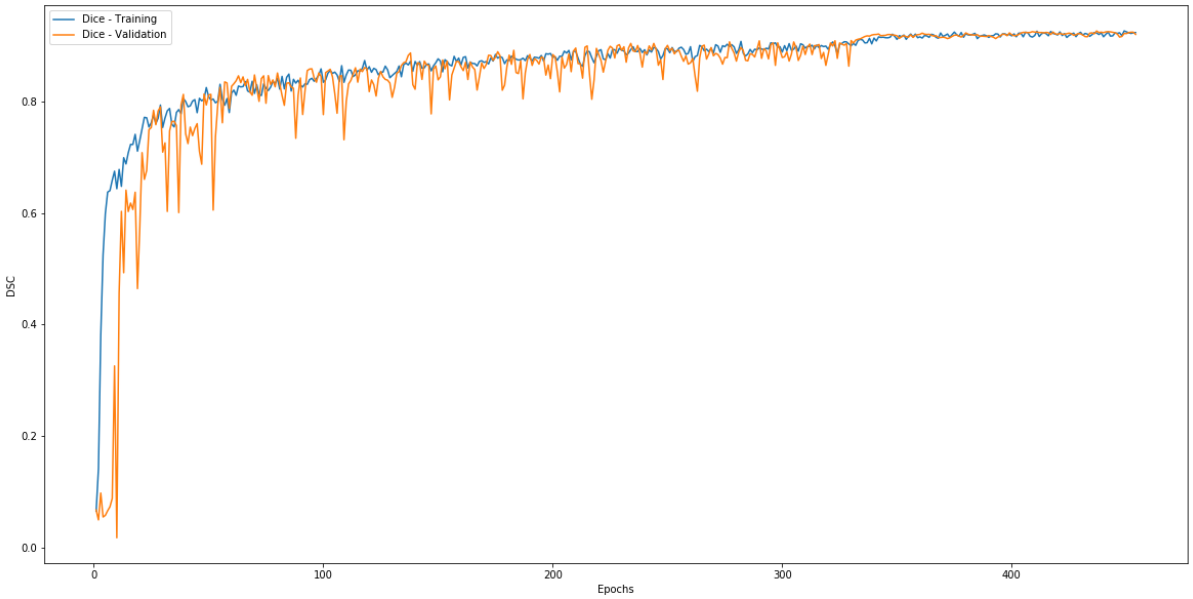

Figure 5.4: Training process of Trial 1 for prostate segmentation. The blue line represents the DSC evolution at the training set and the orange line represents the DSC evolution of the validation set.

Once the model was trained and the weights were stored, comparing the predicted to the manual segmentations was possible. Figure 5.5 describes the twenty best prostate segmentations using Trial 1 weights.

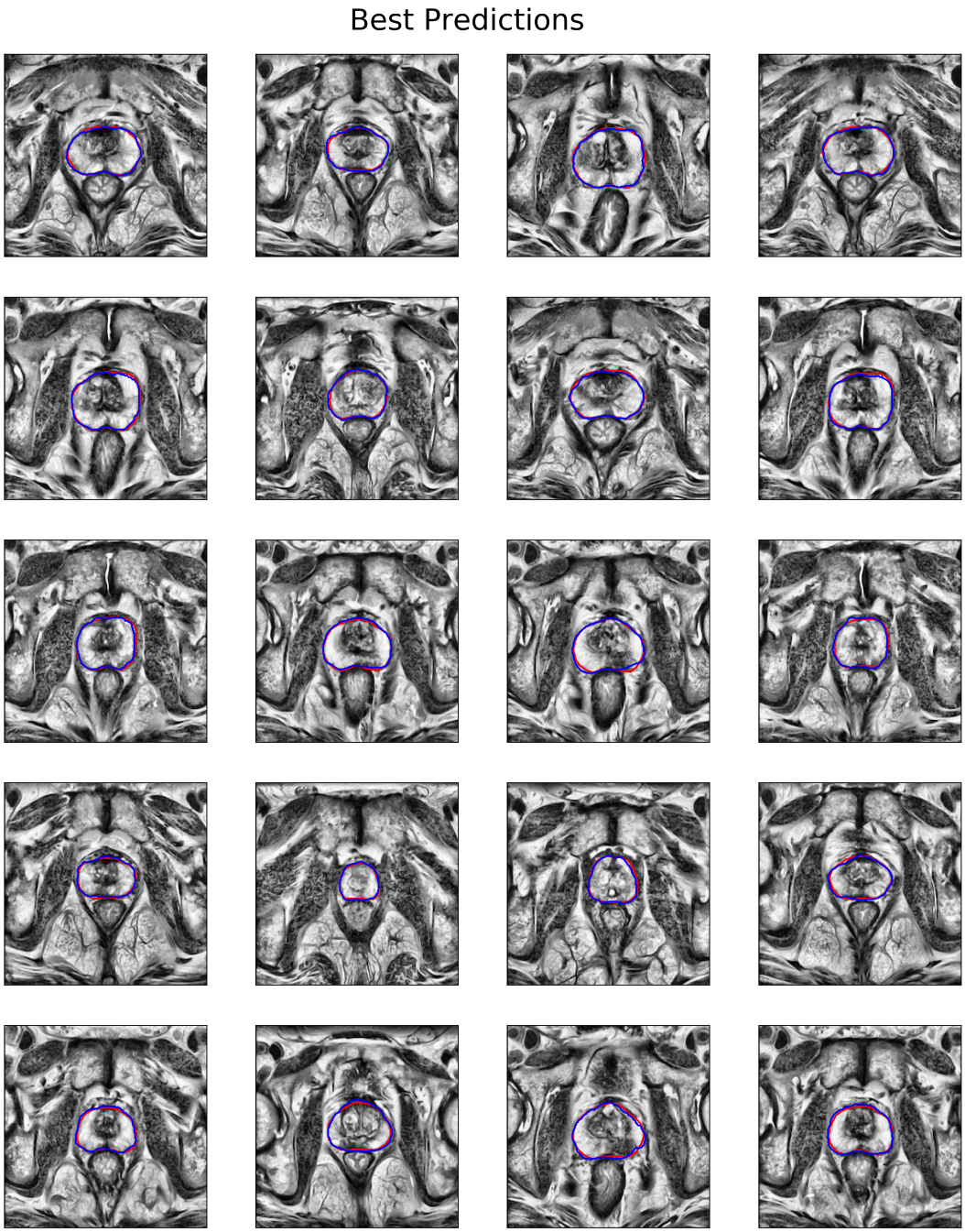

Figure 5.5: Twenty best slice predictions considering Trial 1 weights. The red line represents the manual segmentation contours and the blue line represents the model predicted contours.

The first model was trained, optimised and implemented. Cropping original images and removing the slices where the prostate was not represented was possible, resulting in a ”new” dataset. In Table 5.5, the ”new” dataset details are presented.

Table 5.5: Prostate lesion segmentation datasets comparison.

Prostate Lesion Segmentation		
	Initial Dataset	**Final dataset**
Number of Cases	28	28
Number of Slices	810	370
Image Size	512 x 512	256 x 256

Prostate lesion models were developed considering the best configuration of the previous model. Prostate segmentation Trial 1 yielded the best results, so the subsequent models used the same configuration. The difference between these models was the dataset used, as described in Chapter 4. Models 1 to 4 considered the selected cropped cases, with non prostate slices removed. Conversely, Model 5 used the selected cases with the non-prostate slices removed, but with the original [512 x 512] size. This reference is important to evaluate the role of the zooming process. The models were trained with 4-fold cross-validation and the highest DSC values were extracted. Model 1 results, considering DSC, HD and MSD, are presented in Table 5.6:

Table 5.6: Results of Model 1 cross-validation. * DSC Mean is the highest validation DSC value presented in Model 1 mean training process.

Prostate Lesion Segmentation - Model 1						
	K = 1	**K = 2**	**K = 3**	**K = 4**	**Mean**	**Std Dev**
DSC	0.75	0.73	0.67	0.73	0.71*	0.03
HD	26.3	24.3	25.4	25.2	25.3	0.71
MSD	2.3	2.4	3.1	2.1	2.5	0.38

The following figures represent the training process for each of the 4-folds (Figure 5.6-9) and its mean progress (Figure 5.10):

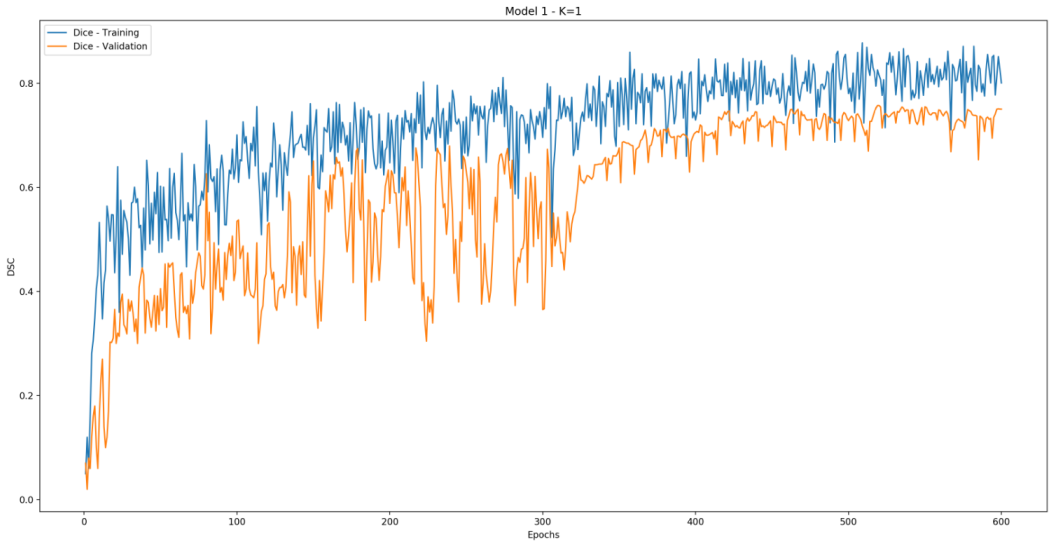

Figure 5.6: Model 1: K = 1 training process.

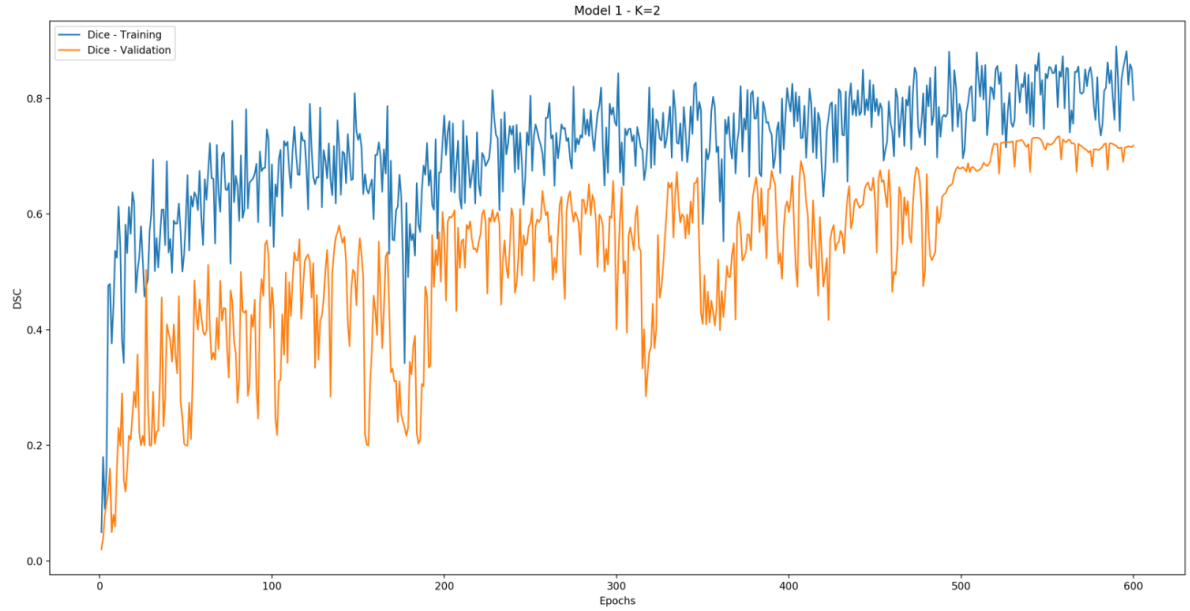

Figure 5.7: Model 1: K = 2 training process.

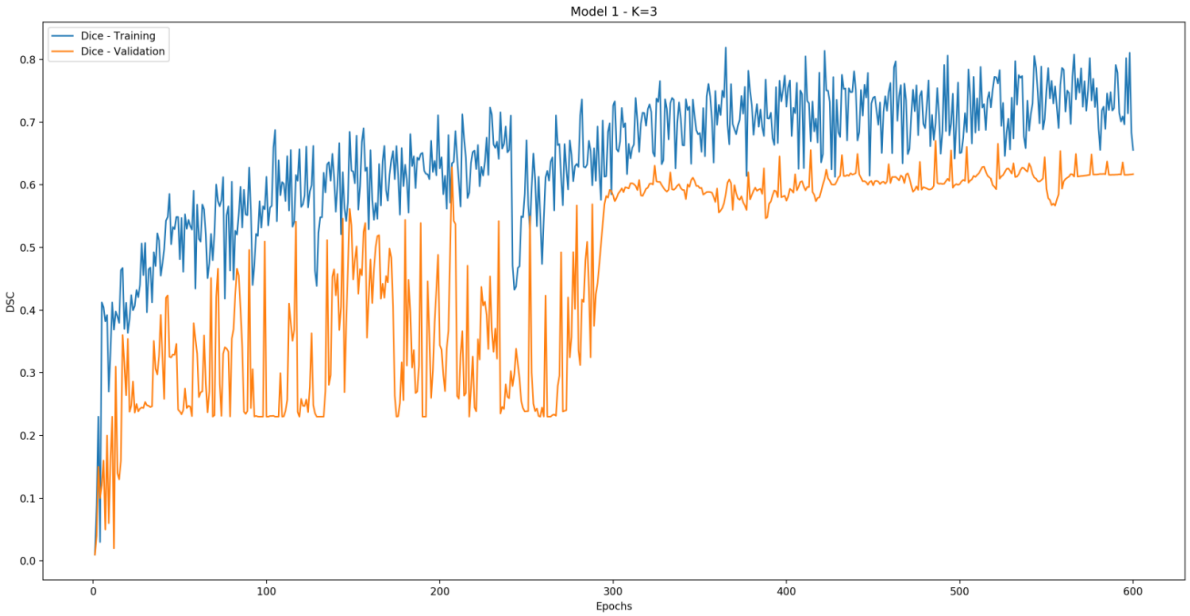

Figure 5.8: Model 1: K = 3 training process.

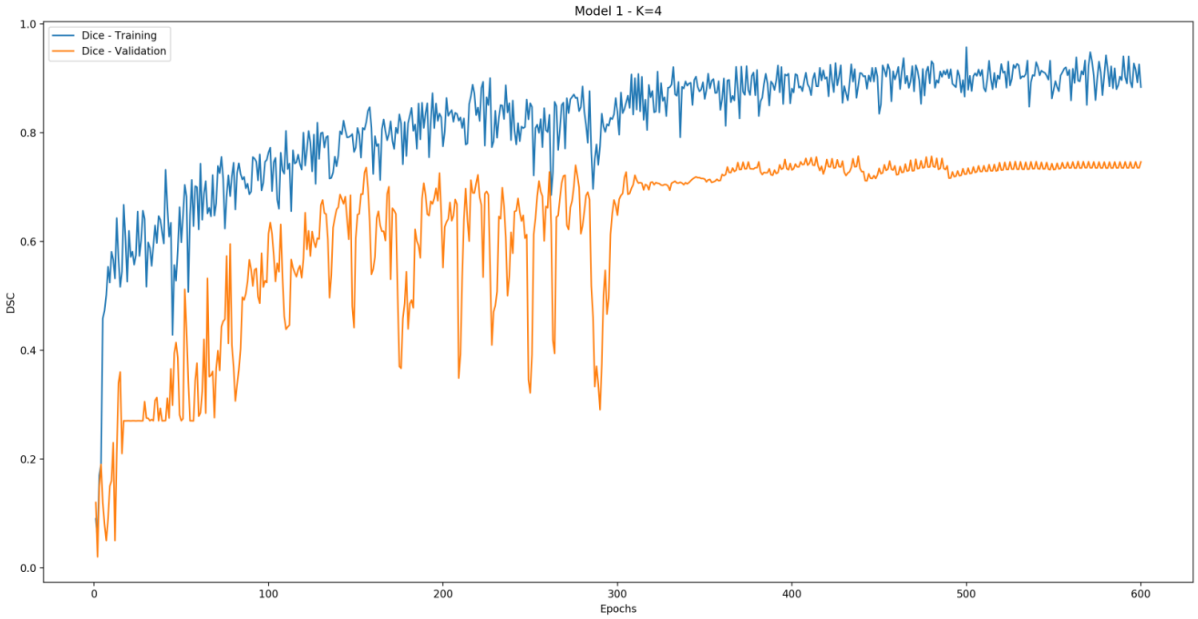

Figure 5.9: Model 1: K = 4 training process.

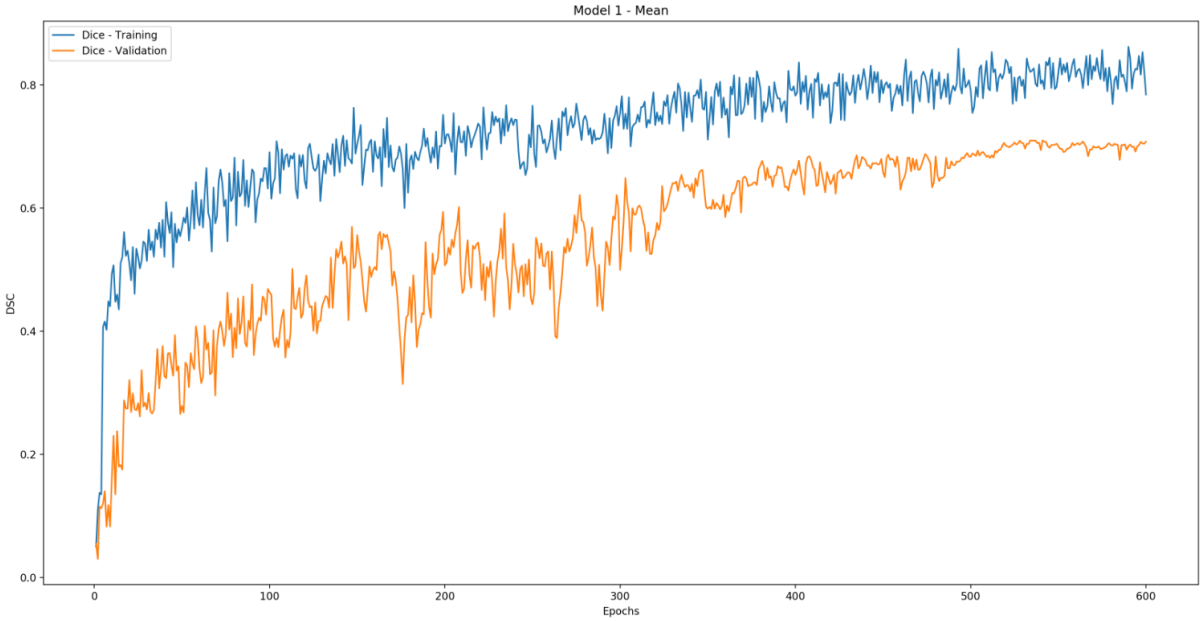

Figure 5.10: Model 1: Mean training process.

The procedures adopted to train Model 2 were identical to Model 1. The following table represents Model 2 results and Figure 5.11 describes Model 2 DSC mean values:

Table 5.7: Results of Model 2 cross-validation.

Prostate Lesion Segmentation - Model 2						
	K = 1	**K = 2**	**K = 3**	**K = 4**	**Mean**	**Std Dev**
DSC	0.71	0.64	0.69	0.72	0.69	0.03
HD	23.3	28.3	23.6	25.2	25.1	1.28
MSD	2.4	3.6	3.1	2.3	2.9	0.53

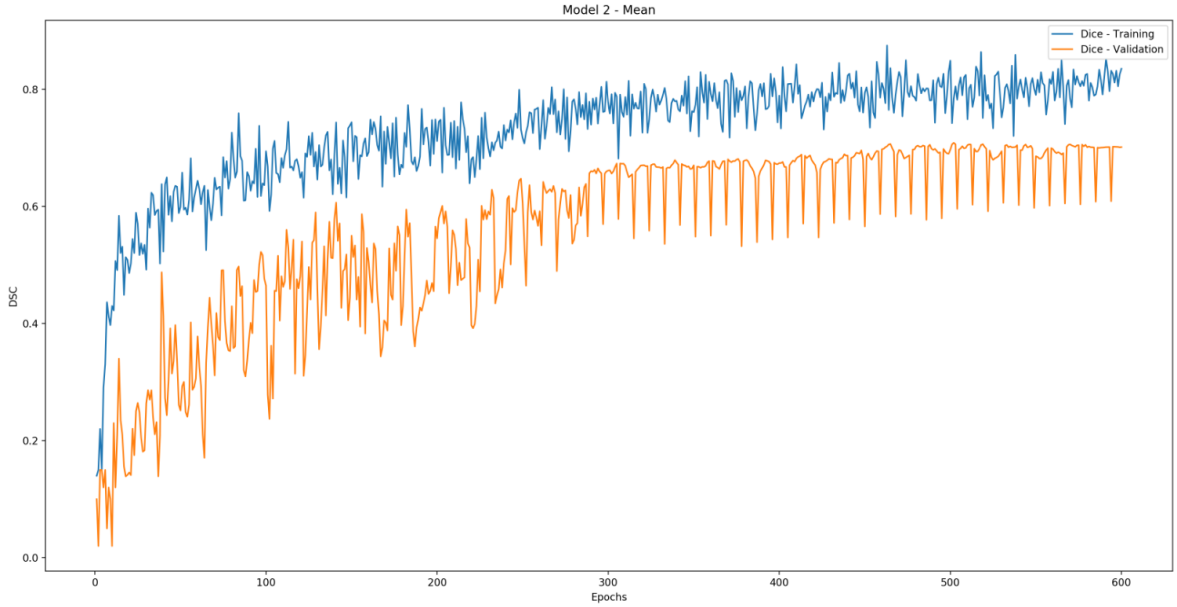

Figure 5.11: Model 2: Mean training process.

Table 5.8 describes Model 3 results and Figure 5.12 its mean DSC progress.

Table 5.8: Results of Model 3 cross-validation.

Prostate Lesion Segmentation - Model 3						
	K = 1	**K = 2**	**K = 3**	**K = 4**	**Mean**	**Std Dev**
DSC	0.76	0.79	0.75	0.79	0.76	0.02
HD	21.1	19.2	20.5	19.8	20.2	0.72
MSD	2.3	1.9	2.4	1.9	2.1	0.23

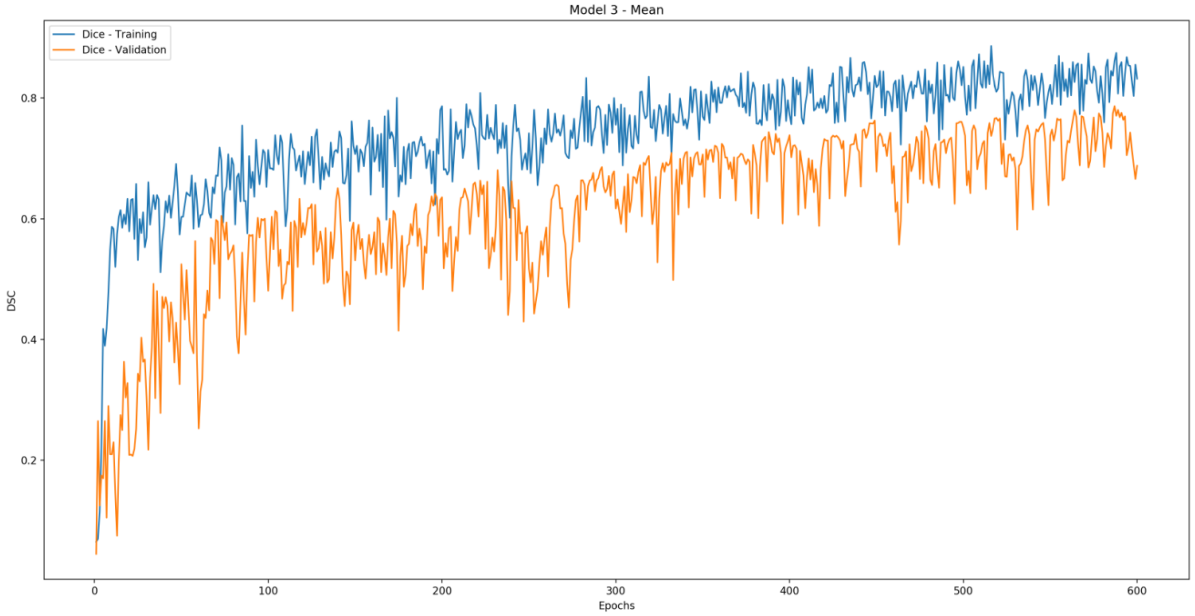

Figure 5.12: Model 3: Mean training process.

The following table describes Model 4 results and Figure 5.13 describes the progress of the mean DSC.

Table 5.9: Results of Model 4 cross-validation.

Prostate Lesion Segmentation - Model 4						
	K = 1	**K = 2**	**K = 3**	**K = 4**	**Mean**	**Std Dev**
DSC	0.76	0.73	0.78	0.70	0.74	0.03
HD	21.4	22.4	21.7	22.6	22.0	0.49
MSD	2.4	2.8	2.2	2.7	2.5	0.24

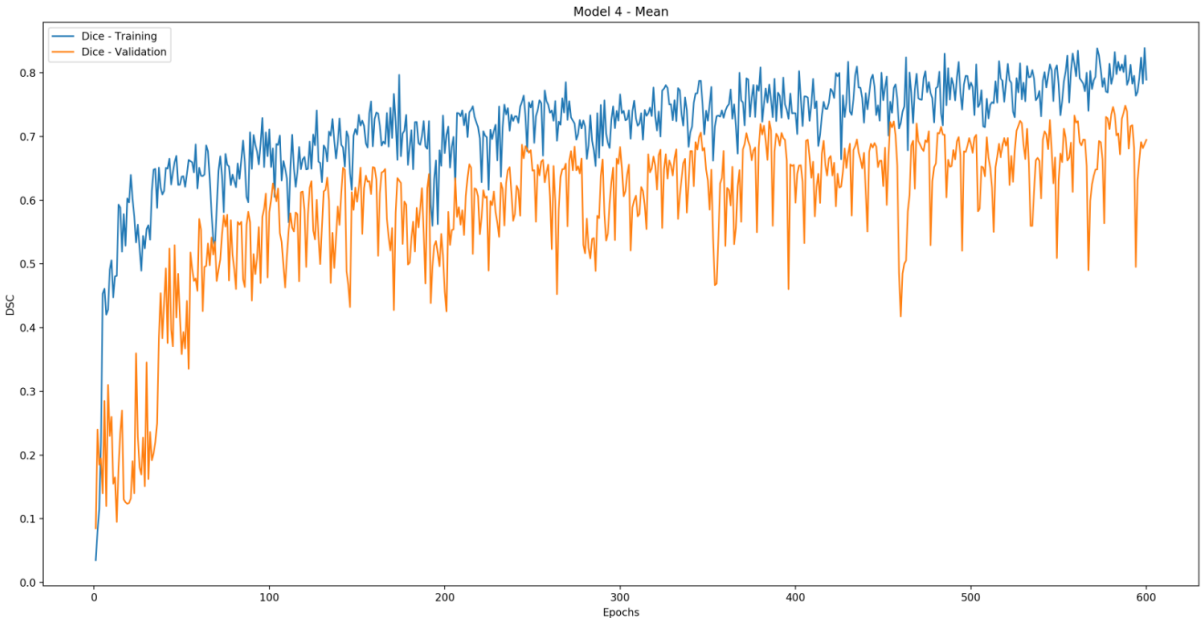

Figure 5.13: Model 4: Mean training process.

The last model, Model 5, the reference network which received non-cropped images was addressed to Table 5.10 and Figure 5.14 presents the mean DSC progress results.

Table 5.10: Results of Model 5 cross-validation.

Prostate Lesion Segmentation - Model 5						
	K = 1	**K = 2**	**K = 3**	**K = 4**	**Mean**	**Std Dev**
DSC	0.51	0.50	0.51	0.54	0.51	0.02
HD	37.5	39.4	33.4	32.1	35.6	2.96
MSD	4.8	5.4	4.9	4.5	4.9	0.32

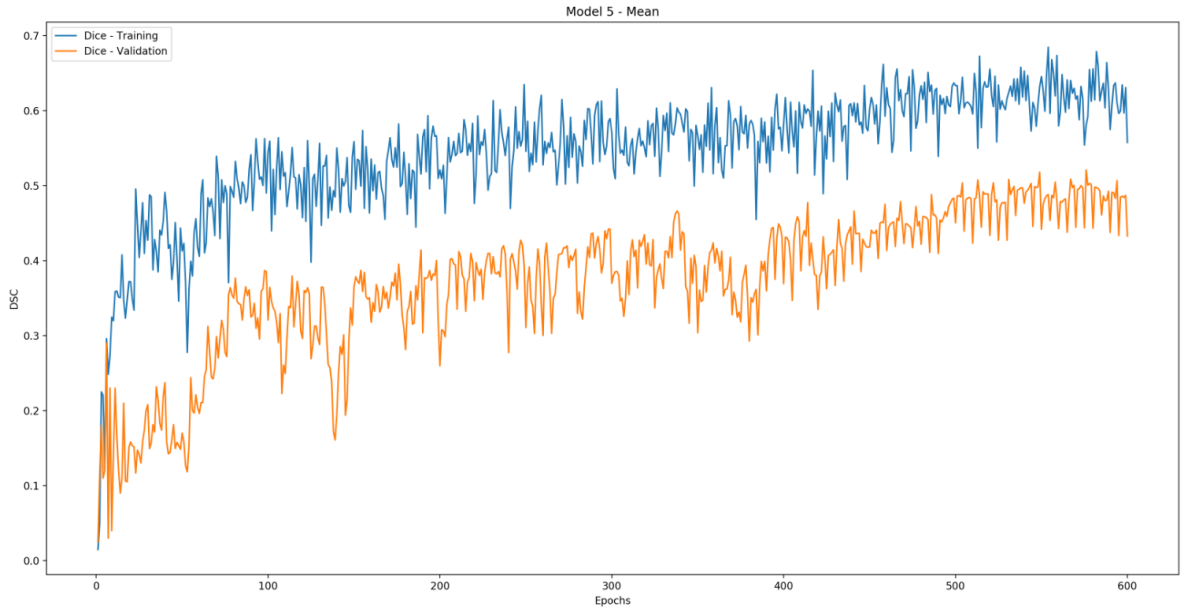

Figure 5.14: Model 5: Mean training process.

Table 5.11 is a summary of prostate lesion segmentation results. It presents the mean DSC, HD and MSD of each model.

Table 5.11: Summary of prostate lesion segmentation results.

Prostate Lesion Segmentation - Summary					
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
DSC	0.71	0.69	0.76	0.74	0.51
HD	25.3	25.1	20.2	22.0	35.6
MSD	2.5	2.9	2.1	2.7	4.9

The five models were trained and implemented. Figure 5.15 and Figure 5.16 represent the predicted lesions contours of the two best models: in T2WIs using Model 3 weights and in ADC images considering Model 4 weights

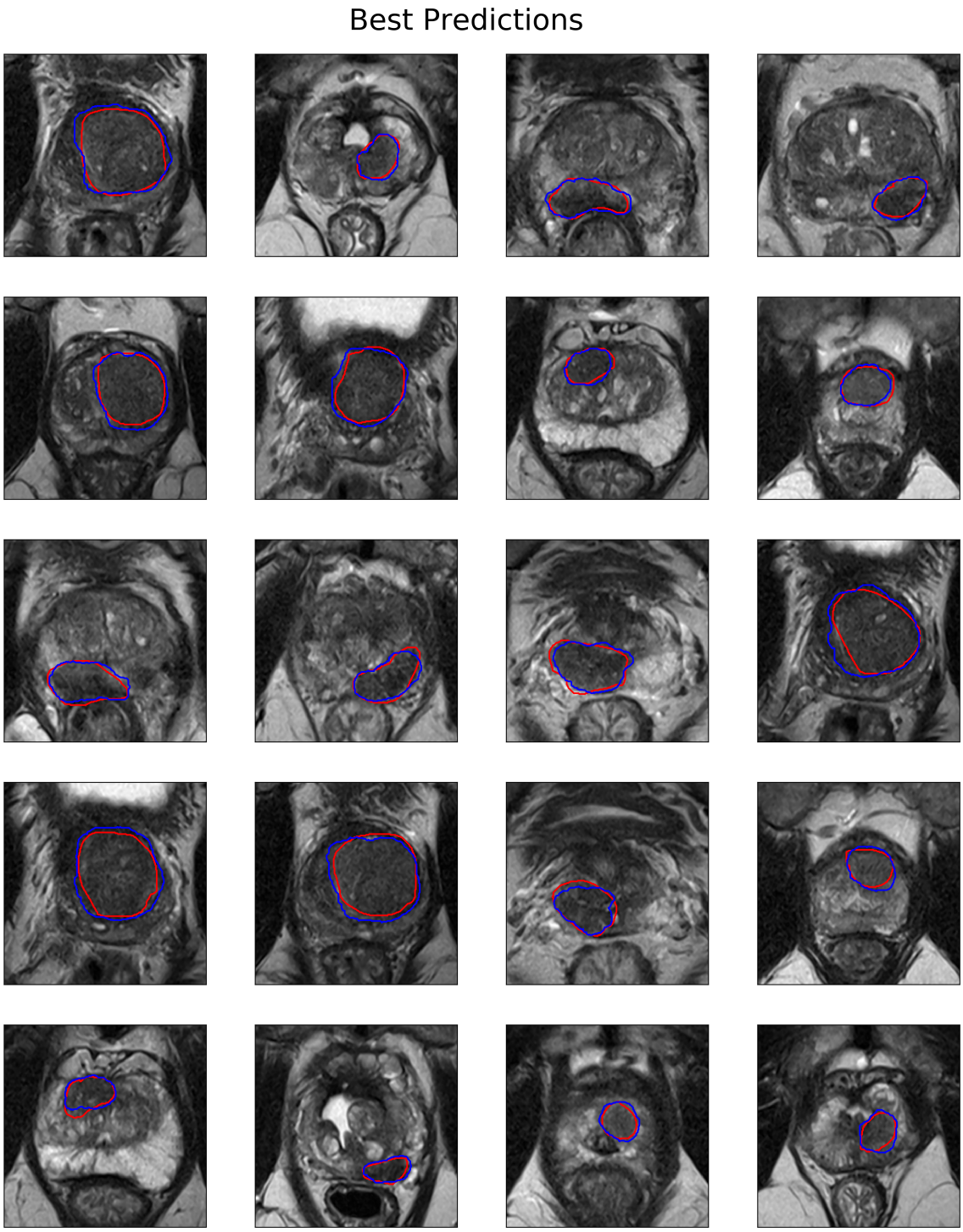

Figure 5.15: Twenty best slice predictions considering Model 3 weights. The red line represents the manual segmentation contours and the blue line indicates the model predicted contours.

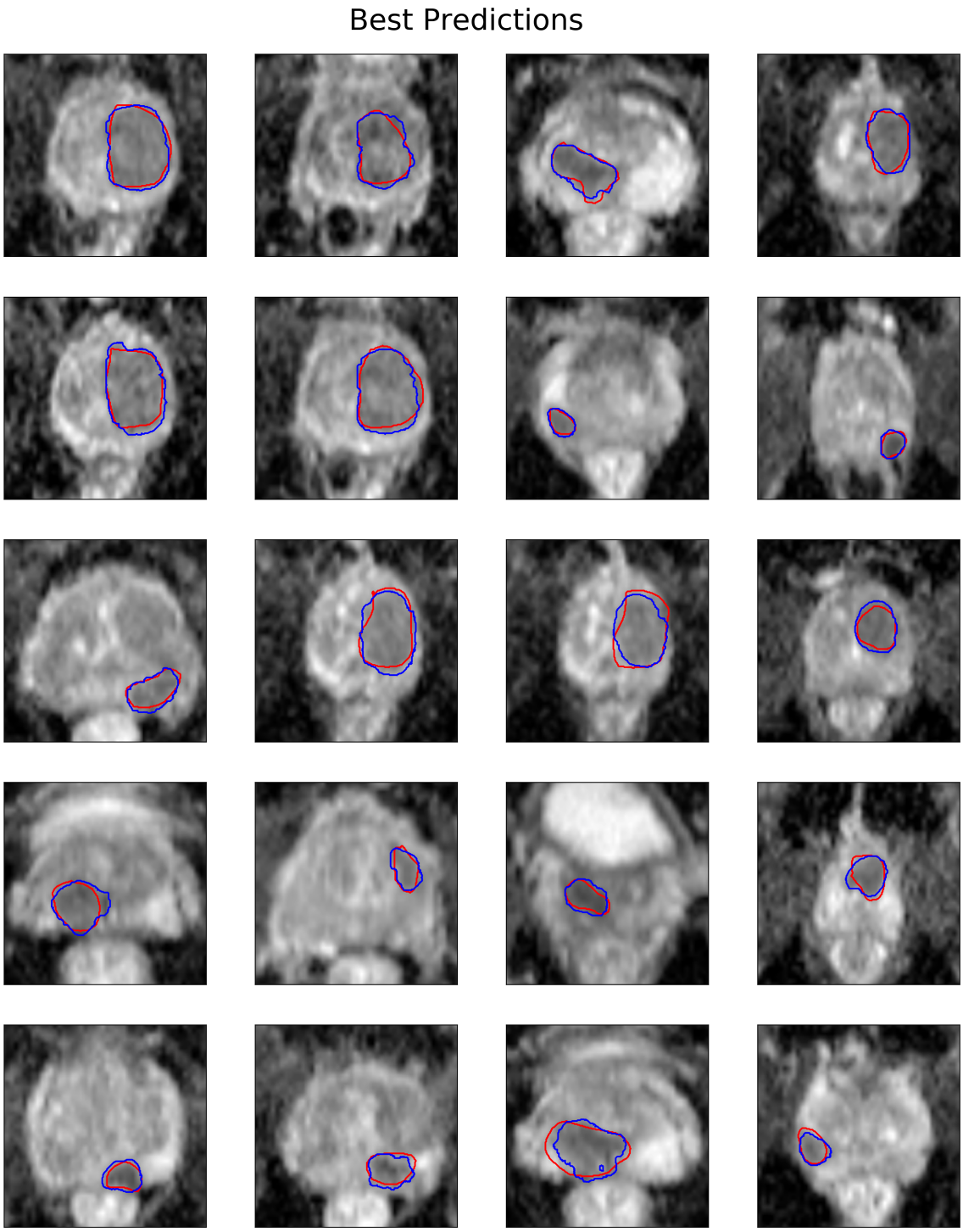

Figure 5.16: Twenty best slice predictions considering Model 4 weights. The red line represents the manual segmentation contours and the blue line indicates the model predicted contours.

6. Discussion

Data collection was the first part of the work towards this dissertation and the masks were created by a radiologist from Hospital da Luz Lisboa, Dr Adalgisa Guerra. The first task was the prostate region segmentation in T2WI. In general, the automatic tool provided by Multi-Parametric Analysis software was assertive and precise regarding the definition of the prostate contours. However, this tool needed to be optimised halfway through the project by Siemens. Until then, the automatic prostate segmentation tool indicated significant difficulties to define the prostate apex limits. Therefore, each slice was manually adjusted to get an optimal segmentation of the prostate zone. With the software update, the problem was solved and the corrective delineations that were needed to perform were residual.

The second task of data collection was the lesion manual segmentations. This part was significantly complex because prostate lesions detection and segmentation is extremely subjective. Pathological anatomy reports were important to know whether there were one or more lesions confirmed in the prostate. The region and classification of those lesions were also present in the report, but not their contours. According to the subjective analysis of this process, detecting the existence of a certain prostate lesion was complicated, even with the pathological anatomy reports. As a result, some manual segmentations may not have been performed correctly. Figure 5.1 illustrates some of the difficulties in segmenting some lesions based on the T2WIs and ADC. Consequently, there were significant limitations in the prostate lesion segmentation models optimisation.

With the data collected, the following step was imaging correction. Several holes were detected in the developed masks, as observed in Figure 5.2, some due to human error during the segmentation process, others due to software limitations which did not fill the delimited area completely. This step was important because if the holes had been kept in the masks, the model would associate that space as non-lesion or non-prostate in the first model. As a result, the holes imaging features and the lesion/prostate imaging features were the same, but presented to the model as different things. These details are relevant because if the data are not well prepared or well acquired, the resulting model will never be acceptable.

The following procedure was the objective evaluation of the data. Intensity pixels histograms were built by comparing the pixel spectrum of the prostate to the pixel spectrum of the lesion. In the case presented in Figure 5.3, there is a considerable difference in the intensity curve shape between the prostate and the lesion. Conversely, in most cases, the curve shape is similar on both sides, being the relevant difference the intensity peak localisation. Considering the fact that most cases presented curves with similar shapes and close peaks, the difficulty that models could have during the training process is enhanced. As described in Chapter 5, the mean difference between the intensity peaks when considering all dataset is only 67, the maximum difference 243 and the lowest value is 3. These values are important to explain the difficulty that health professionals have in detecting or segmenting prostate lesions. The lesions are

usually not clear and well defined to be completely identified by humans and, consequently, an automatic process could be the solution or, at least, a significant help.

The last step of data processing consisted of data selection. This process was in part performed subjectively, considering the data sensitivity acquired during this dissertation, but also taking into account the results of intensity pixels histograms. If a lesion was not easily delineated by human eye and also presented identical histograms of the prostate and itself, with the same curve shape and peak localisation, the respective case was discarded regarding prostate lesions segmentation model. This happened because some lesions were not detected during manual segmentation and the masks were performed considering the pathological anatomy reports. Cases where the lesion was impossible to detect by the radiologist were not considered in the model. Therefore, 28 cases with 810 slices were selected in an optimum state to be applied in the model: visible, or at least detectable, by human eye or/and with histograms with different shapes and peaks localisation.

The prostate region model was implemented using the original and pre-selected 53 cases, with 1534 slices. This model purpose was to identify the prostate contours and create a mask of the region. As a result, it would be possible to crop the original images, according to the prostate size plus a safety margin, and with that provide cropped images of the region of interest to the lesion models.

At first, when models were being trained with CPU, the time length of each epoch was about 1 minute and a half. Considering the cross-validation process with 4-folds, the entire model training time length for lesion segmentation, would take more than 60 hours. With GPU application, each epoch took 12 seconds to be performed, which totals of 8 hours to train a single model. This is a considerable difference and the reason why GPU was selected to train deep neural networks. This highlights the impact of GPU applications in medical imaging applications.

Several trials were completed for prostate segmentation and the three best are described in Table 5.3. The trial which yielded the best results was Trial 1 with initial LR of 0.01, dropout of 50%, batch size of 32 and architecture depth of 4. Trial 2 had five blocks of depth and a batch size of 16. Compared to the best trial, one more block of depth means more parameters to train and optimise. Usually, with the residual blocks that were applied, more depth would mean better results. However, considering the number of epochs that was defined (600 epochs) for prostate segmentation, Trial 2 had more difficulties to optimise to the same level as Trial 1 in the same time-space. Conversely, Trial 3 started with an initial LR that was too high and the model could not converge to a minimum value of DSC loss. Although *Reduce LR On Plateau* was applied, there were not enough epochs to match Trial 1 results. In all prostate models, the *Early Stopping* command was called after 125 epochs without validation DSC improvement. Trial 1, for example, stopped after 464 epochs. As observed in Table 5.4, Trial 1 DSC was registered with 0.88 for the validation set. It is likely that with more data and more precise segmentations, the results could have been better. In terms of HD and MSD, the results were also similar to the state of the art references. Trial 1 has shown a MSD of 2.1 mm, which is a small value for this model purpose. As a result, the model was considered sufficiently good to be used to crop the original images.

Trial 1 values for prostate segmentation are equivalent to state-of-the-art results, especially the results from the Department of Medical Physics and Biomedical Engineering team of University College of London, described in Table 3.5. In Table 3.5, U-Net presented a mean DSC of 0.84, while the network

with the highest value, the HighRes3dNet, presented a value of 0.88. In the PROMISE 12 challenge, the best network for prostate segmentation was the Imorphics, as described in Table 3.3, with a DSC of 0.879. Considering these state-of-the-art values and the DSC as the main metric for segmentation, the results acquired in this section of the dissertation with Trial 1 are considered promising.

The selected prostate segmentation model was applied to the initial dataset to crop the original images. As described in Table 5.5, the image size was reduced from [512 x 512] to [256 x 256]. In this reduced size, the prostate zone represents more than 90% of the image, while with the original size, the prostate zone was approximately 50% of the entire image. The number of slices was also reduced in order to focus exclusively on slices where the prostate had been identified. As observed in Table 5.5, prostate slices represent around 50% of the volumes and 370 slices were considered to be used in prostate lesion segmentation. Cropping the images could be important to focus the network into imaging variability inside the prostate and the cutting slices process would be relevant to significantly reduce the following models training time.

As referred, five models were trained to segment prostate lesions. The first two models received only one image as input (T2 or ADC), while Model 3 and Model 4 received the two types of images as input. Model 5 is the reference to evaluate the results of cropping the original images. Table 5.6 presents a 4-fold cross-validation of Model 1 with DSC values around 0.70 for the validation set and 0.80 for the training set. With every fold DSC arrays, the mean values of those arrays combined were plotted as ”Model 1: Mean training process”, Figure 5.10. The best epoch - i.e. the epoch with the highest validation DSC value - achieved a 0.71 value. This is considered the mean DSC value of Model 1. The other metrics, HD and MSD, are the average of the 4-folds performance. This model has reached 25.3 mm in terms of HD and 2.5 mm considering MSD.

Additionally, as described in Table 5.7, Model 2 presented a mean DSC of 0.69, HD of 25.1 mm and MSD of 2.9. Comparing to Model 1, the results are similar. Considering T2WI, defining the contours of a visible lesion is easier than contouring it in ADC, while in ADC images it is easier to detect the existence of a lesion. Since this is a segmentation project and lesions delineation is the main task, the previous fact about the features of both types of images can explain the slight difference in favour of Model 1.

Model 3 was the model with the best results, as presented in Table 5.8. It revealed a mean DSC value of 0.76, 20.2 mm of HD and 2.1 mm of MSD. Both ADC and T2WI served as input, while the masks were designed in T2WIs. Masks designed with T2 images, as previously described, are better in terms of contouring the lesions. In addition, lesion information resulting from T2 and ADC images is considered in this model. The output of the model is a single mask and the input are the original T2 and ADC images. Consequently, information of both type of images is grouped and taken into account during model training. This result enhances the fact that both images have complementary information that can be useful for prostate lesion segmentation.

Moreover, Model 4 also presented reasonable results. Mean DSC of 0.74, HD of 22.0 mm and MSD of 2.7 mm are the outcomes of the model. The DSC value is lower than Model 3, but it is higher than the models with simply one original image as input. This statement contributes to the importance of combining both imaging features in the same model.

In the end, Model 5 revealed the lowest mean DSC with 0.51, as referred to in Table 5.11. This value is considerably lower compared to other models and this might be explained because there are more parts of the image to process that are not relevant. Considering just one step to segment directly prostate lesions - i.e. without the prostate region model and consequent imaging crop - the slices where the prostate is not present remain in dataset volumes, providing, once again, the model with more irrelevant information. Combining these facts, the difficulty of training a model with assertiveness and precision in these circumstances is enhanced. HD and MSD values also corroborate that statement.

In terms of state-of-the-art results for prostate lesion segmentation, a research team recently developed a radiomics deeply supervised U-Net for automated detection of the prostate [57]. The developed project has similarities with this dissertation and presented a DSC of 0.91 with T2WI, exclusively. It is an extremely good performance, but it is important to mention that the this team incorporated radiomic texture features within the CNN that they selected. Moreover, a research team from China also revealed great results when fusing information from T2WIs with ADC images [58]. They developed a model to detect prostate lesions, instead of segmenting them, and information from both sequences was considered. Therefore, combining networks which analyse texture features with models that receive two images at a time, from different sequences, could be the following step to improve this dissertation results.

As a conclusion, models which received more information as input (Model 3 and Model 4) presented better results than the models with simply one image type as input. Combining information in segmentation projects is important, because more useful information indicates more chances of identifying the proper contours. Another conclusion is the fact that prostate lesion segmentation is significantly better when performed in two steps - prostate model segmentation to crop the region of interest first - than simply with one step - directly identifying the lesions contours from raw images. However, the absolute values of lesion segmentation DSCs did not perform as well as when compared to prostate segmentation or other segmentation projects. Training a model to segment artefacts that even health professionals have difficulties to identify is not trivial, but we hypothesize that with more quality data, the baseline of these results would be improved.

7. Conclusion and Future Work

Prostate cancer is one of the most common types of cancer, affecting millions of men, every year. In order to improve the diagnosis processes of this disease, the present work successfully adapted the U-Net architecture to consider the automatic segmentation of prostate lesions in magnetic resonant images.

This dissertation presented five models to segment prostate lesions and the difference between them was the dataset used. Model 1 used T2 original images with lesions masks created in T2WIs. Model 2 considered ADC original images and lesions masks were produced in ADC images. Conversely, both Model 3 and Model 4 received T2 and ADC original images at once as input, with lesion masks created in T2 or ADC, respectively. In the end, Model 5 was implemented to be used as a reference and to evaluate the potential of previously cropping original images. The data type used to train this model is the same as Model 1, however, the image size was not reduced and slices without prostate were also included.

Model 3 yielded the best results, with a mean DSC of 0.76, HD of 20.2 mm and MSD of 2.1 mm. Both Model 3 and Model 4 had better results than the models that only used one image type as input (Model 1 and Model 2). This is important to emphasise the impact of combining multiparametric information in the same network. In addition, Model 5 presented the worst results, highlighting the positive impact of cropping original images before trying to segment prostate lesions. In this way, it is possible to conclude that multiparametric information is better than just considering individual sequences and that enlarging the region of interest is beneficial for prostate lesion segmentation.

In conclusion, models which received T2 and ADC images as input presented better results than the models with one image type as input. Another statement is the fact that prostate lesion segmentation has shown better results when performed in two steps - i.e. prostate model segmentation with imaging isolation - than simply with one step - i.e. directly identifying the lesions contours from raw images.

For future work, some limitations of the presented study should be addressed, in particular, the quantity and quality of manual segmentations. Even for a trained professional, identifying the contours of a prostate lesion with total certainty is extremely difficult. This is a limitation that could be overtaken by having more than one radiologist creating the masks, promoting the necessary data variability. With just one professional segmenting lesions, it is not possible to evaluate inter-user variability. Ideally, several radiologists should create the masks and compare them. Building deeper networks can also be important to improve the actual results, however, the equipment used to train these CNNs should have specific GPUs, in order to reduce the training period significantly. In the future, this project can also be evaluated and implemented considering the number of detected lesions, instead of simply contemplating the DSC.

Finally, bearing in mind all of the above mentioned, it is possible to conclude that the proposed goal for this project was achieved. A model to segment prostate lesions was created and with reasonable results. So far, this model cannot be used independently to identify prostate lesions, but can be important to reduce

the exam analysis time by giving healthcare professionals a validation role. Even if the segmentations are not completely correct, the model could be used to warn the professionals to the possible existence of a lesion in a certain location. In the end, this dissertation is simply one more scientific step to improve healthcare diagnosis strategies, combining the potential of engineering with medicine.

References

[1] Amis, E. and Lang, E., 1994. Radiology Of The Lower Urinary Tract. Berlin: Springer-Verlag. vii, 5, 6

[2] L. V. Kost and G. W. Evans, "Occurrence and significance of striated muscle within the prostate.," The Journal of Urology, vol. 92, pp. 703–4, 1964. 5

[3] P. Dasgupta, ABC of Prostate Cancer (ABC Series Book 193). BMJ Books, 2011. vii, 6, 8

[4] A. A. Villers, J. E. McNeal, E. A. Redwine, F. S. Freiha, and T. A. Stamey, "Pathogenesis and biological significance of seminal vesicle invasion in prostatic adenocarcinoma.," The Journal of Urology, vol. 143, no. 6, pp. 1183–7, 199. 5

[5] M. Ohori, P. T. Scardino, S. L. Lapin, C. Seale-Hawkins, J. Link, and T. M. Wheeler, "The mechanisms and prognostic significance of seminal vesicle involvement by prostate cancer.," The American Journal of Surgical Pathology, vol. 17, no. 12, pp. 1252–61, 1993. 5

[6] A. G. Ayala, J. Y. Ro, R. Babaian, P. Troncoso, and D. J. Grignon, "The prostatic capsule: does it exist? Its importance in the staging and treatment of prostatic carcinoma.," The American Journal of Surgical Pathology, vol. 13, no. 1, pp. 21–7, 1989. 6

[7] A. Lopez-Beltran, L. Cheng, R. Montironi, and M. R. Raspollini, Pathology of the Prostate. Cambridge University Press, 2017. 6

[8] Gco.iarc.fr. (2019). Global Cancer Observatory (Globocan). [online] Available at: https://gco.iarc.fr. 7

[9] Apurologia.pt. (2019). Associação Portuguesa de Urologia. [online] Available at: https://www.apurologia.pt/publico/cancro_da_prostata.htm. 7

[10] Salinas C., Tsodikov A. et Ishak-Howard M. Prostate Cancer in Young Men: An Important Clinical Enity (2014). Public Access, 11(1), 1–23. 7

[11] G. Ploussard, J. I. Epstein, R. Montironi, P. R. Carroll, M. Wirth, M.-O. Grimm, A. S. Bjartell, F. Montorsi, S. J. Freedland, A. Erbersdobler, and T. H. van der Kwast, "The Contemporary Concept of Significant Versus Insignificant Prostate Cancer.," European Urology, vol. 60, no. 2, pp. 291–303, 2011. 7

[12] T. Wolters, M. J. Roobol, P. J. Van Leeuwen, R. C. N. Van Den Bergh, R. F. Hoedemaeker, G. J. L. H. Van Leenders, F. H. Schr¨oder, and T. H. Van Der Kwast, "A Critical Analysis of the Tumor Volume Threshold for Clinically Insignificant Prostate Cancer Using a Data Set of a Randomized Screening Trial.," The Journal of Urology, 2011. 7

[13] W. Hamilton and D. Sharp, "Symptomatic diagnosis of prostate cancer in primary care: a structured review.," The British Journal of General Practice : the Journal of the Royal College of General Practitioners, vol. 54, no. 505, pp. 617–21, 2004. 7

[14] S.-M. Young, P. Bansal, E. T. Vella, A. Finelli, C. Levitt, and A. Loblaw, "Systematic review of clinical features of suspected prostate cancer in primary care.," College of Family Physicians of Canada, vol. 61, no. 1, pp. e26–35, 2015. 7

[15] J. V. Tricoli, M. Schoenfeldt, and B. A. Conley, "Detection of Prostate Cancer and Predicting Progression: Current and Future Diagnostic Markers.," tech. rep., 2004. 7

[16] M. J. Barry and L. H. Simmons, "Prevention of Prostate Cancer Morbidity and Mortality: Primary Prevention and Early Detection.," The Medical Clinics of North America, vol. 101, no. 4, pp. 787–806, 2017. 7

[17] Mayoclinic.org. (2019). Prostate cancer - Diagnosis and treatment - Mayo Clinic. [online] Available at: https://www.mayoclinic.org/diseases-conditions/prostatecancer/diagnosis-treatment/drc-20353093. 8

[18] Marcus, D. M., Rossi, P. J., Nour, S. G., Jani, A. B. (2014). The impact of multiparametric pelvic magnetic resonance imaging on risk stratification in patients with localised prostate cancer. Urology, 84(1), 132–137. 8

[19] Marks, L., Kamrava, M., Kishan A. U., Margolis D. J (2016). Multiparametric MRI for prostate cancer improves Gleason score assessment in favorable risk prostate cancer. 8

[20] N. Mottet, J. Bellmunt, M. Bolla, E. Briers, M. G. Cumberbatch, M. De Santis, N. Fossati, T. Gross, A. M. Henry, S. Joniau, T. B. Lam, M. D. Mason, V. B. Matveev, P. C. Moldovan, R. C. van den Bergh, T. Van den Broeck, H. G. van der Poel, T. H. van der Kwast, O. Rouvi‘ere, I. G. Schoots, T. Wiegel, and P. Cornford, "EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent.," European Urology, vol. 71, no. 4, pp. 618–629, 2017. 8

[21] W. B. Roberts and M. Han, "Clinical significance and treatment of biochemical recurrence after definitive therapy for localized prostate cancer.," Surgical Oncology, vol. 18, no. 3, pp. 268–274, 2009. 8

[22] S. Sridharan, V. Macias, K. Tangella, J. Melamed, E. Dube, M. X. Kong, A. Kajdacsy-Balla, and G. Popescu, "Prediction of prostate cancer recurrence using quantitative phase imaging: Validation on a general population.," Scientific Reports, vol. 6, pp. 1–10, 2016. 8

[23] J. L. Stanford, A. S. Hamilton, F. D. Gilliland, R. A. Stephenson, J. W. Eley, P. C. Albertsen, L. C. Harlan, and A. L. Potosky, "After Radical Prostatectomy for Clinically Localized Prostate Cancer The Prostate Cancer Outcomes Study.," vol. 1024, 2018. 8

[24] Y. Du, Q. Long, B. Guan, L. Mu, J. Tian, Y. Jiang, X. Bai, and D. Wu, "Robot-Assisted Radical Prostatectomy Is More Beneficial for Prostate Cancer Patients: A System Review and Meta-Analysis.," Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, vol. 24, pp. 272–287, 2018. 8

[25] Huynh, L. M. and Ahlering, T. E. (2018). Robot-Assisted Radical Prostatectomy: A Step-by-Step Guide. Journal of Endourology, 32(May), S28–S32. vii, 9

[26] S. F. Shariat, M. W. Kattan, A. J. Vickers, P. I. Karakiewicz, and P. T. Scardino, "Critical review of prostate cancer predictive tools," Future Oncology, vol. 5, no. 10, pp. 1555–1584, 2009. 8

[27] J. F. Ward, M. L. Blute, J. Slezak, E. J. Bergstralh, and H. Zincke, "The long-term clinical impact of biochemical recurrence of prostate cancer 5 or more years after radical prostatectomy.," Journal of Urology, vol. 170, no. 5, pp. 1872–1876, 2003. 9

[28] G. W. Hull, F. Rabbani, F. Abbas, T. M. Wheeler, M. W. Kattan, and P. T. Scardino, "Cancer control with radical prostatectomy alone in 1,000 consecutive patients.," The Journal of Urology, vol. 167, no. 2 Pt 1, pp. 528–34, 2002. 9

[29] M. Moschini, V. Sharma, F. Zattoni, J. F. Quevedo, B. J. Davis, E. Kwon, and R. J. Karnes, "Natural History of Clinical Recurrence Patterns of Lymph Node–Positive Prostate Cancer After Radical Prostatectomy.," European Urology, vol. 69, no. 1, pp. 135–142, 2016. 9

[30] Chollet, F., 2017. Deep Learning With Python. 1st ed. Manning Publications. 10

[31] Litjens, G., Toth, R., Ven, W. Van De, Hoeks, C., Kerkstra, S., Malmberg, F., Kirschner, M. (2015). Evaluation of prostate segmentation algorithms for MRI: the PROMISE12. Med Image Anal, 18(2), 359–373. 10

[32] Jung, S. K., Kim, T. W. (2016). New approach for the diagnosis of extractions with neural network machine learning. American Journal of Orthodontics and Dentofacial Orthopedics, pp. 127–133. vii, 11, 14

[33] Dev.to. (2019). Creating of neural network using JavaScript - DEV Community [online] Available at: https://dev.to/liashchynskyi/creating-of-neural-network-usingjavascript-in-7minutes-o21. vii, 12

[34] Online course: A. Ng (2019). Deep Learning Specialization Program. Neural Networks and Deep Learning. Class: Activation functions. https://www.deeplearning.ai/deeplearning-specialization/ 12

[35] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, no. 6088, pp. 533–536, 1986. 12

[36] Online course: A. Ng (2019). Deep Learning Specialization Program. Improving Deep Neural Networks: Hyperparameter tuning, Regularization and optimization: Class: Adam optimisation algorithm. https://www.deeplearning.ai/deep-learningspecialization/ 13

[37] Online course: A. Ng (2019). Deep Learning Specialization Program. Improving Deep Neural Networks: Hyperparameter tuning, Regularization and optimisation: Class: Regularization. https://www.deeplearning.ai/deep-learning-specialization/ 14

[38] Online course: A. Ng (2019). Deep Learning Specialization Program. Improving Deep Neural Networks: Hyperparameter tuning, Regularization and optimisation: Class: Dropout regularization. https://www.deeplearning.ai/deep-learning-specialization/ 14

[39] Online course: A. Ng (2019). Deep Learning Specialization Program. Improving Deep Neural Networks: Hyperparameter tuning, Regularization and optimisation: Class: Basic recipe for machine learning. https://www.deeplearning.ai/deep-learningspecialization/ 15

[40] Online course: A. Ng (2019). Deep Learning Specialization Program. Improving Deep Neural Networks: Hyperparameter tuning, Regularization and optimisation: Class: Learning rate decay. https://www.deeplearning.ai/deep-learning-specialization/ 15

[41] Ghavami, N., Hu, Y., Gibson, E., Bonmati, E., Emberton, M., Moore, C. M., Barratt, D. C. (2019). Automatic segmentation of prostate MRI using convolutional neural networks: Investigating the impact of network architecture on the accuracy of volume measurement and MRI-ultrasound registration. Medical Image Analysis. vii, ix, 16, 21, 22, 23, 25, 26

[42] Milletari, F., Navab, N., Ahmadi, S. (2016). V-Net: Fully Convolutional Neural Networks, pp. 1–11. vii, ix, 16, 21, 22, 24, 25

[43] Karimi, D., Salcudean, S. E. (2019). Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks. IEEE Transactions on Medical Imaging. vii, 16

[44] Abdulkadir, A., Lienkamp, S. S., Brox, T., Ronneberger, O. (n.d.). 3D U-Net: Learning Dense Volumetric. 424–432. vii, ix, 16, 21, 24

[45] Litjens, G., Toth, R., Ven, W. Van De, Hoeks, C., Kerkstra, S., Malmberg, F.,Kirschner, M. (2015). Evaluation of prostate segmentation algorithms for MRI: the PROMISE12. Med Image Anal, 18(2), 359–373. 16

[46] Medium. (2019). A Comprehensive Guide to Convolutional Neural Networks—theELI5 way. [online] Available at: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. vii, 17

[47] Online course: A. Ng (2019). Deep Learning Specialization Program. Convolutional Neural Networks course. Class: Padding. https://www.deeplearning.ai/deep-learning-specialization 18

[48] Online course: A. Ng (2019). Deep Learning Specialization Program. Convolutional Neural Networks course. Class: Strided convolutions. https://www.deeplearning.ai/deep-learning-specialization 18

[49] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D. (1989). Backpropagation applied to handwritten zip code recognition. Neural Computation 1(4), 541–551. 19

[50] Krizhevsky, A., Sutskever, I., Hinton, G.E.(2012). Imagenet classification with deep convolutional neural networks. 19

[51] Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. 19

[52] Seyedhosseini, M., Sajjadi, M., Tasdizen, T. (2013). Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 2168–2175. 19

[53] Hariharan, B., Arbelaez, P., Girshick, R., Malik, J.. (2014) Hypercolumns for object segmentation and fine-grained localisation. 19

[54] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Heidelberg. vii, ix, 19, 20, 21, 25, 30

[55] Ghavami, N., Hu, Y., Bonmati, E., Gibson, E., Ghavami, N., Hu, Y., Barratt, D. (2018). Automatic slice segmentation of intraoperative transrectal ultrasound images using convolutional neural networks. vii, 21, 22, 24, 25

[56] Lovegrove, C. E., Matanhelia, M., Randeva, J., Eldred Evans, D., Tam, H., Miah, S., Shah, T. T. (2018). Prostate imaging features that indicate benign or Malignant pathology on biopsy. Translational Andrology and Urology, 7 (Suppl 4), S420–S435. 26

[57] Praful Hambarde, Sanjay Talbar, Abhishek Mahajan, Satishkumar Chavan, Meenakshi Thakur, Nilesh Sable, Prostate lesion segmentation in MR images using radiomics based deeply supervised U-Net, Biocybernetics and Biomedical Engineering, Volume 40, Issue 4, 2020, Pages 1421-1435. 52

[58] Kwang-Ting (Tim) Cheng, Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI, Medical Image Analysis, Volume 42, 2017, Pages 212-227. 52